

COVER PAGE

Award Number: W81XWH-14-1-0224

TITLE: Environmental Mycobiome Modifiers of Inflammation and Fibrosis in Systemic Sclerosis

PRINCIPAL INVESTIGATOR: Michael L. Whitfield, Ph.D

CONTRACTING ORGANIZATION: The Trustees of Dartmouth College
Hanover, NH 03755

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE		<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.			
1. REPORT DATE October 2015	2. REPORT TYPE Annual	3. DATES COVERED 20 Aug 2014-19 Aug 2015	
4. TITLE AND SUBTITLE Environmental Mycobiome Modifiers of Inflammation and Fibrosis in Systemic Sclerosis		5a. CONTRACT NUMBER	
		5b. GRANT NUMBER W81XWH-14-1-0224	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Michael Whitfield, Ph.D, Patricia A. Pioli, PhD. Robert Lafyatis, MD Sarah Arron, MD/Ph.D E-Mail: Michael.L.Whitfield@dartmouth.edu		5d. PROJECT NUMBER	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Geisel School of Medicine at Dartmouth, Hanover, NH, Boston University Medical Center, MA, University of California, San Francisco, CA		8. PERFORMING ORGANIZATION REPORT	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSOR/MONITOR'S ACRONYM(S)	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT This project is focused on Systemic Sclerosis (SSc), a progressive fibrotic disease characterized by skin fibrosis and damage to internal organs. While a wide range of environmental and biological triggers have been proposed, no definitive etiologic agents have yet been identified. Metagenomic analysis of non-human sequences in SSc RNA-seq data was used to detect microbial sequences in human tissues in an unbiased, quantitative manner. Our studies suggest that disease pathogenesis includes a common environmental fungal trigger, <i>Rhodotorula glutinis</i> , which we hypothesize elicits immune activation in a permissive host genetic background. Skin biopsies have been collected from SSc patients and analyzed by high-throughput sequencing, providing substantial gene expression data as well as detailed information regarding the host microbiome. Data has been compared against that of healthy control samples.			

15. SUBJECT TERMS: IMSA, systemic sclerosis, scleroderma, mycobiome, fibrosis, gene, genetics, RNA-seq, <i>R. glutinis</i> , Metagenomics					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF	18. # OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	54	19b. TELEPHONE NUMBER (include area code)

Table of Contents

1. INTRODUCTION	5
2. KEYWORDS:	5
3. ACCOMPLISHMENTS	5
Milestone 1	5
Milestone 2	5
Milestone 3	6
PRELIMINARY RESULTS BY MILESTONE	6
KEY RESEARCH ACCOMPLISHMENTS Summary (Jul 1, 2011-June 30, 2014)	12
The next reporting period:	12
4. IMPACT	12
5. CHANGES/PROBLEMS	12
6. PRODUCTS:	12
Oral Presentations: (Chronological Order)	12
Abstracts and Presentations: (Chronological Order)	13
Manuscripts:	13
Degrees obtained that are supported by this award	15
Development of cell lines, tissue or serum repositories	15
7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS	15
8. SPECIAL REPORTING REQUIREMENTS	15
9. REFERENCES	15
10. APPENDIX	16

1. INTRODUCTION

Systemic sclerosis (SSc) is a heterogeneous disease of fibrosis and inflammation, concomitant with significant autoimmunity. SSc often presents with skin manifestations and Raynaud's phenomenon; the extent and location of fibrotic lesions in people with SSc contributes to the diagnoses of disease subtypes and prognosis. My laboratory has pioneered the use of gene expression subsets in SSc [1-4]. Most recently we have demonstrated enrichment of a mycobiome component (*Rhodotorula glutinis*) in SSc patient skin [5].

We describe our studies from the first year of the grant below. This work was accomplished by researchers at Geisel School of Medicine at Dartmouth, Boston University Medical Center and University of California, San Francisco under the partnering PI option.

2. KEYWORDS:

IMSA, systemic sclerosis, scleroderma, SSc, mycobiome, microbiome, fibrosis, gene, genetics, RNA-seq, Next Generation Sequencing, skin, *R. glutinis*, *Rhodotorula*, Metagenomics,

3. ACCOMPLISHMENTS

Milestones were assigned to this proposal, with tasks to be accomplished by each investigator. The overall **summary** of our progress relative to these tasks is given below, followed by a complete discussion of our work this past year.

Milestone 1 Determine the identity and distribution of microbiome components across SSc skin.

Task 1 (Months 1-36) Whitfield Laboratory to perform RNA-seq analysis of SSc skin biopsies.

Including technical replicates, RNA-seq has been run on 18 SSc patient skin biopsies to date. Recruitment of additional SSc patients and healthy controls is ongoing.

Task 2 (Months 6-36) Whitfield Laboratory to perform RNA-seq analysis for differentially expressed mRNAs and non-coding RNAs.

Raw sequence reads have been analyzed using publicly available software packages that have been optimized and validated by us.

Task 3 (Months 6-36) Arron group to perform IMSA and determine the identity of microbiome components.

Dr. Arron's group is currently performing metagenomic analysis on 35 new and existing skin biopsy samples (31 SSc and 4 healthy controls) from the Whitfield Laboratory.

Task 4 (Months 1-24) Arron group to create scaffolds from aligned reads for each microbiome component and develop nested PCR followed by targeted multiplexed sequencing assays for cost-effective screening.

We have developed a fast and efficient nested PCR reaction targeting microbiome components specific for fungal species identification. We then evaluated the identity and number of fungal reads by next generation sequencing (see Task 5).

Task 5 (Months 1-12) Whitfield Laboratory to examine a larger population of archived skin biopsy RNA to determine the prevalence of microbiome components across the SSc population.

Initial analyses of microbiome components within archived RNA samples were performed using both nested PCR and NanoString-based methods. This will be ongoing in year 2.

Task 6 (Months 1-24) Culture microbiome components from the skin of SSc patients. Use of skin biopsies as a method for fungal culture was not successful.

*Currently skin swabs are being used as a means of microbial collection prior to biopsy. *Rhodotorula* spp., as well as a variety of other yeast and molds, have been isolated from control patients. Expansion of patient swabbing efforts as a means of fungal detection is ongoing.*

Milestone 2 Identify the inflammatory infiltrates in SSc skin and their response to microbiome components

Task 1 (Months 1-6) Whitfield Laboratory to perform computational analysis/prediction of inflammatory cell infiltrates from whole genome expression data.

We used single sample Gene Set Enrichment Analysis (ssGSEA) to identify the cellular subsets in SSc skin at different stages of disease.

Task 2 (Months 6-24) Lafyatis' group to perform immunohistochemistry to validate the computational predictions of task 1 above.

We are currently optimizing markers for different cell types in SSc skin. We will use CD163 for macrophages, and CD3 for T cells. The precise series of markers is ongoing and being carefully defined.

Task 3 (Months 1-18) Whitfield Laboratory to develop protocols for the isolation and characterization of immune cells from skin using the sclerodermatous Graft-Versus Host Disease (sclGVHD) mouse including detailed characterization of cell types.

Institutional approvals have been finalized. We have moved the establishment of the sclGVHD model to year 2. Cell isolation procedures will be tested in 10 – 20 mice.

Task 4 (Months 6-18) Identify the secreted mediators of fibrosis / inflammation being produced (Whitfield / Pioli). Once cells are isolated, we will screen for secreted pro-fibrotic mediators.

Work underway.

Task 5 (Months 12-36) Apply protocols to characterize the inflammatory infiltrate in the skin of SSc patients (Whitfield / Pioli). After cell isolation procedures have been optimized in the sclGVHD mouse we will examine the infiltrate and profibrotic mediators in SSc skin biopsies.

Work underway.

Milestone 3 Determine if SSc patients have a specific immune response against *R. glutinis* that is different from healthy controls and if this response can drive fibrosis.

Task 1 (Months 1-24) Test patient sera for cross-reactivity against *R. glutinis* antigens (Whitfield/Lafyatis).

We have performed western blots using whole cell lysates and probed with sera collected from both healthy controls and SSc patients.

Task 2 (Months 1-24) Identify the cross-reacting proteins by mass spectrometry (Whitfield).

*Serum-immunoprecipitation of *R. glutinis* and human HeLa cell whole cell lysates followed by mass spectrometry was performed to identify immunoreactive proteins associated with *R. glutinis*. We have written a manuscript on the human cross-reactivity. We are having difficulty with the annotation state of the *R. glutinis* genome for annotating those spectra.*

Task 3 (Months 12-36) Use isolated PBMCs and isolated monocytes to examine the cytokines secreted and changes in gene expression when cells are exposed to *R. glutinis* or other putative micro / mycobiome triggers (Whitfield/Pioli).

Work underway.

Task 4 (Months 12-24) Determine if chronic exposure to *R. glutinis* or other micro / mycobiome components stimulate a fibrotic response in a mouse model of SSc. (Whitfield).

Work underway.

PRELIMINARY RESULTS BY MILESTONE

Milestone 1: Determine the identity and distribution of microbiome components across SSc skin

Task 1: RNA-Seq analysis of SSc skin. Including technical replicates, RNA-seq has been run on 18 SSc patient skin biopsies to date. Recruitment of additional SSc patients and healthy controls is ongoing. mRNA from 18 SSc patients have been sequenced, yielding 30-237 million paired-end reads per sample. Reads were

then aligned to the human genome (hg19 assembly). Approximately 70%-80% of reads were uniquely mapped, which is in line with expectations (Table 1).

Table 1. Statistics of alignment

Sample	Read number	Read length	Mapped unique	Mapped length	Mapped mutli	Unmapped multi+	Unmapped short	Unmapped other
MW7017_S2	236958774	150	76.40%	145.2	9.50%	0.00%	13.90%	0.20%
JPO4-FA_S4	128473304	148	80.40%	144	7.30%	0.00%	12.00%	0.30%
KLO6-B_S5	125597225	148	80.60%	143.6	7.80%	0.10%	11.30%	0.30%
MW7021_S4	56139654	149	77.20%	145.2	8.80%	0.10%	13.60%	0.30%
MW7018_S3	39702156	149	78.40%	145.3	6.50%	0.00%	14.90%	0.20%
MW7022_S5	74556388	150	78.80%	146.2	8.00%	0.10%	12.90%	0.10%
MW7015_S1	69172362	149	71.30%	144.3	15.10%	0.00%	13.40%	0.20%
N1_Base	42555825	102	77.30%	99.3	4.60%	0.00%	17.90%	0.10%
N10_Base	44359538	102	81.20%	100	4.20%	0.00%	14.50%	0.10%
N11_Base	32353212	102	80.50%	99.9	4.00%	0.00%	15.40%	0.10%
N18_Base	43103004	102	79.90%	99.9	3.90%	0.00%	16.00%	0.10%
N5_Base	35722430	102	78.70%	99.8	4.20%	0.00%	17.00%	0.10%
N15_Base	30947648	102	80.20%	99.9	4.10%	0.00%	15.60%	0.10%
N7_Base	39145387	102	80.00%	100	4.00%	0.00%	15.80%	0.10%
N9_Base	35556626	102	78.50%	99.9	4.10%	0.00%	17.20%	0.10%
AMO2-FA_S1	113089971	148	78.20%	143.7	8.00%	0.00%	13.60%	0.20%
KBO3-FA_S2	85152721	147	79.20%	143.6	6.70%	0.00%	13.70%	0.50%
KBO3-B_S3	112943946	148	80.30%	143.8	7.80%	0.10%	11.60%	0.20%

Task 2 RNA-seq analysis for differentially expressed mRNAs. In order to identify differentially expressed mRNA, we used RSEM software to estimate the abundance of each mRNA transcript. Each sample was normalized using quantile normalization. Batch biases generated by the inclusion of previously sequenced samples from a separate study (N_Base samples) was performed with ComBat (Figure 1).

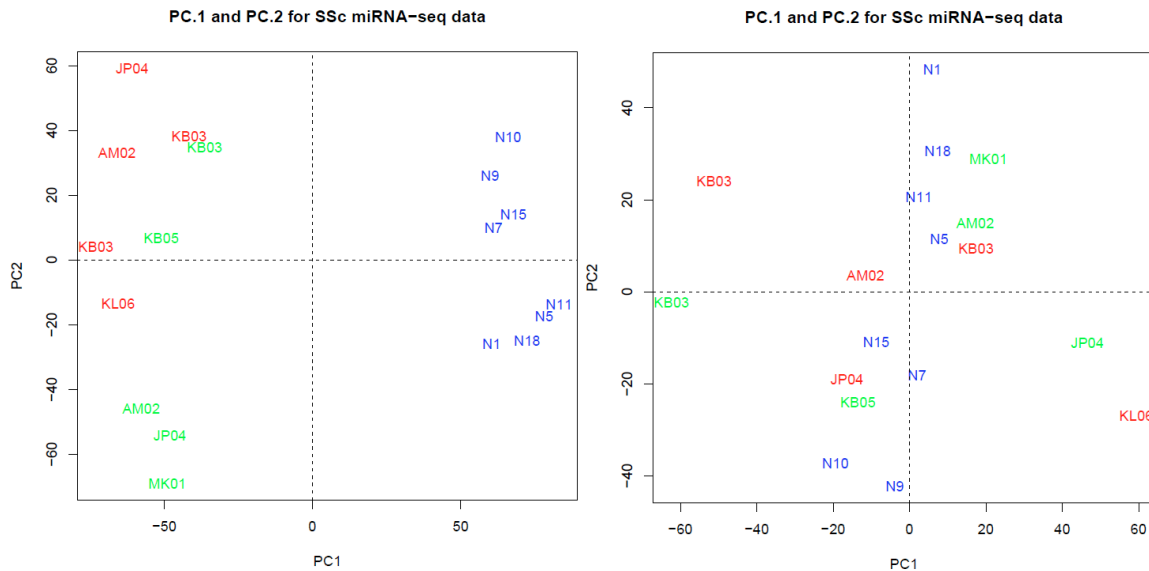
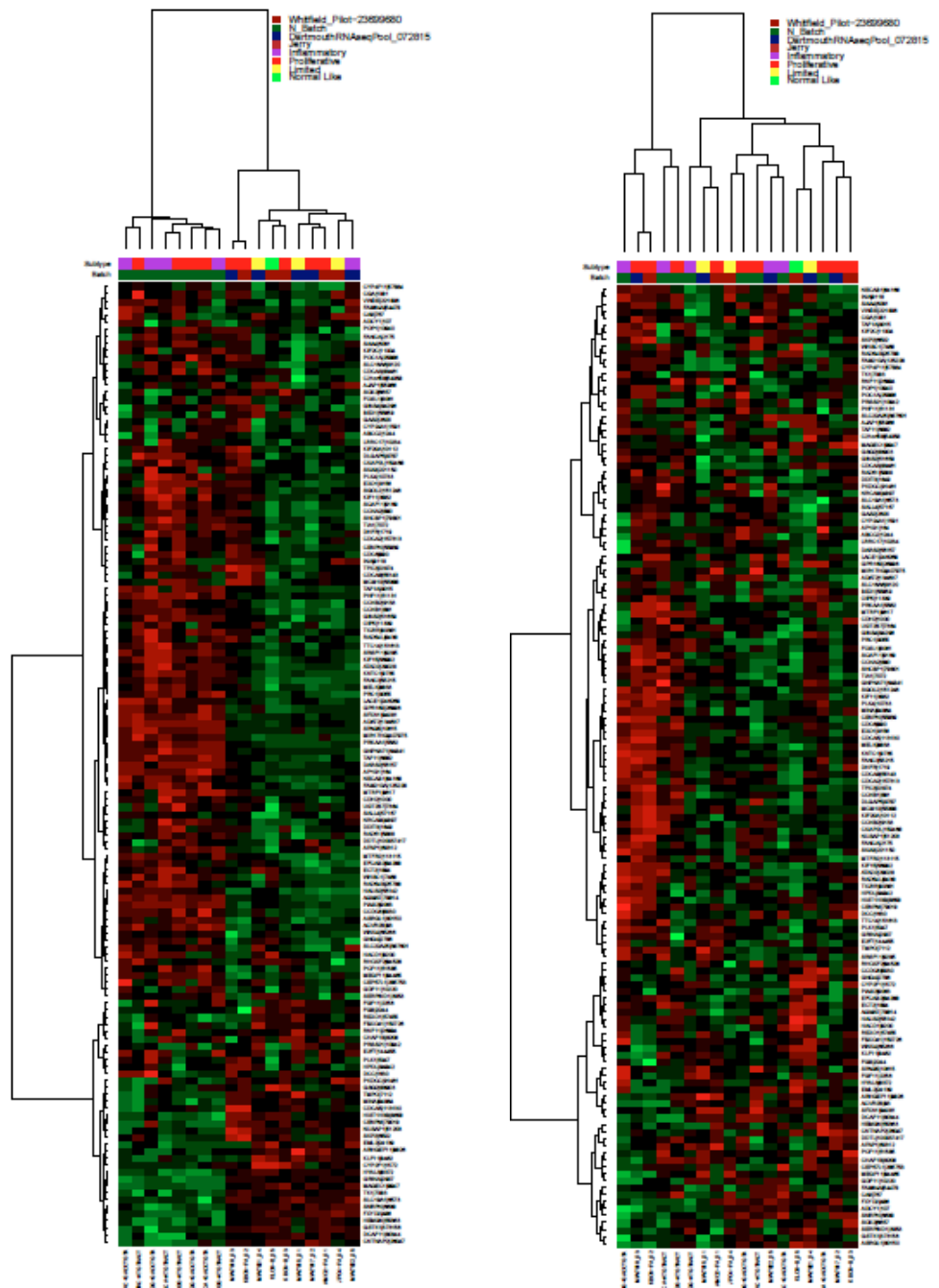


Figure 1.
PCA plot of
measurements. (Left)
no ComBat
correction;
(Right) with
ComBat
correction.

As an initial analysis of these data we chose to examine the consensus genes from Mahoney et al. [4]. These are genes that were consistently and reproducibly associated with individual SSc intrinsic gene expression subsets across three independent patient cohorts. Expression of these genes in our RNA-seq data reveals increased expression in the inflammatory and fibroproliferative subsets of patients (Figure 2). Expression of

these genes is shown both before and after batch correction. Intermixing of samples is clearly evident after ComBat correction, indicating that batch correction was successful.

Figure 2. Heatmap of Matt_267modules. (Left) no ComBat correction; (Right) with ComBat correction.



Task 3: IMSA analysis to identity microbiome components. In order to map the microbiome components present in these skin biopsies, IMSA has been performed on RNA-seq data from both new and previously analyzed ([5]; Li et al. *In preparation*) SSc skin biopsy samples (31 SSc and 4 healthy controls). We performed quality filtering and human sequence filtering using human genome (hg19). Over 99% of the total readset was derived from human or nonhuman primates in both SSc and control samples. IMSA was used to map reads to the NCBI non-redundant nucleotide (nt) database and generate taxonomy reports. In this analysis, each taxonomic level is given a score based on the number of reads aligning to sequences in that taxonomic category, where reads with multiple best alignments generate partial scores for each category with an alignment. From preliminary data analysis, we find that only inflammatory samples have high *Rhodotorula glutinis* target read counts (Figure 3) and the lowest species diversity (Figure 4), consistent with the preliminary data we presented in our initial grant proposal. Therefore, this preliminary analysis validates those original data.

Figure 3. IMSA analysis of RNA-seq data from SSc skin biopsies. SSc skin biopsies were divided by intrinsic gene expression subset, as previously described [1, 4]. Each biopsy was analyzed for *R. glutinis* sequences and IMSA score plotted for each subset.

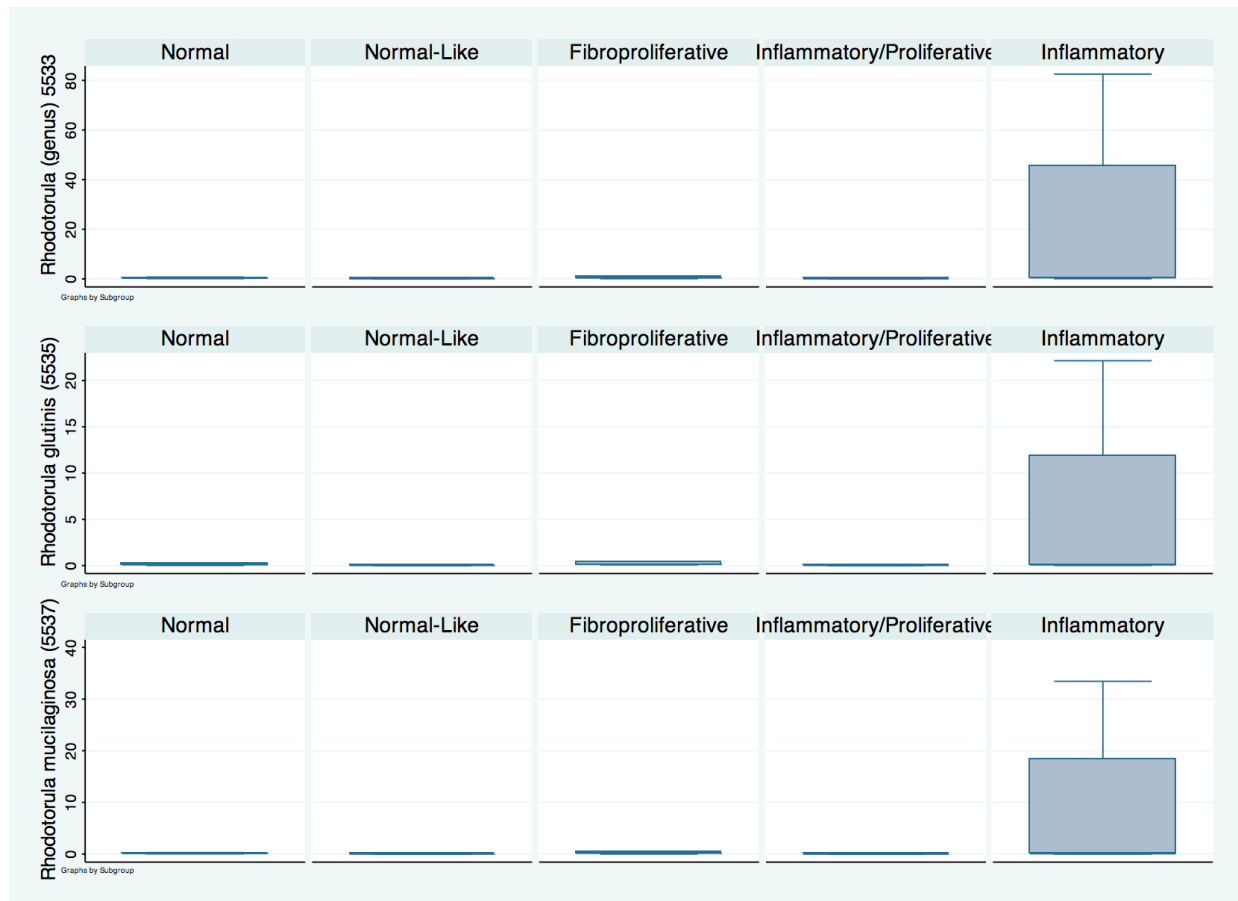
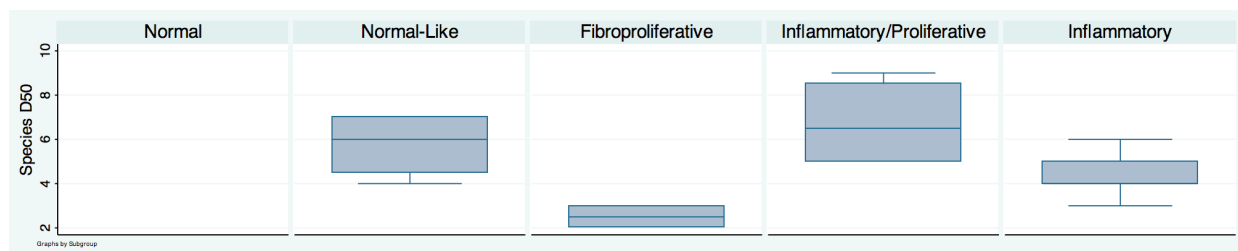


Figure 4. Measure of species diversity from metagenomic analysis of SSc skin biopsies. Using IMSA scores, we find the greatest non-human species diversity in the normal-like and inflammatory-proliferative intrinsic subsets. The lowest species diversity is present in the inflammatory subset, which reflects the increase in *R. glutinis* reads.



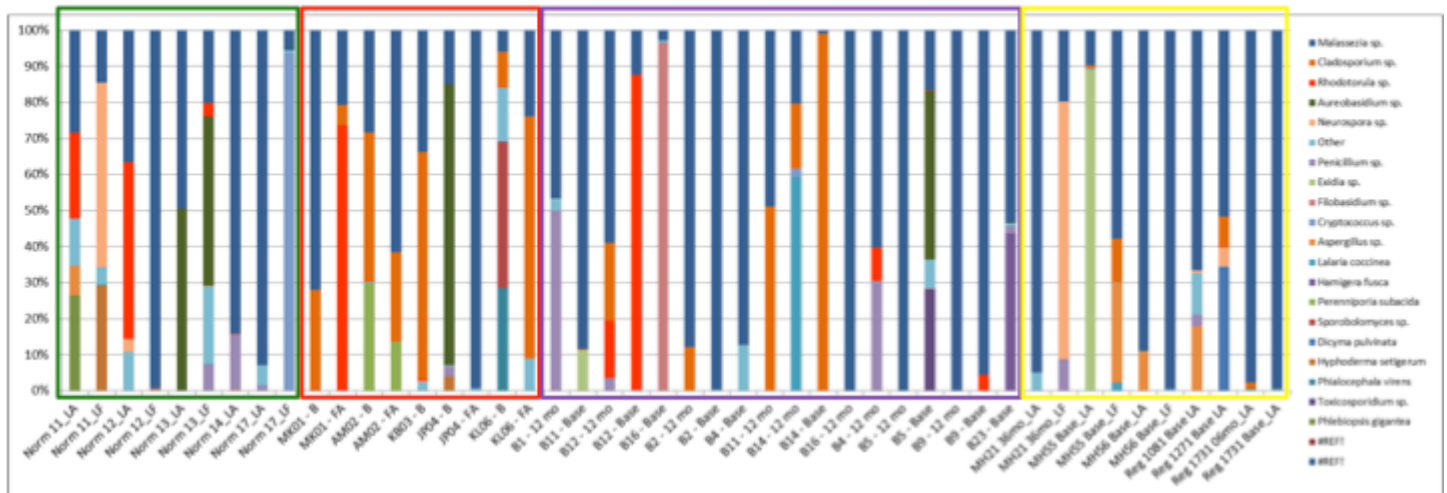
Tasks 4 and 5. Develop a nested PCR-based assay followed by targeted multiplexed sequencing as a cost-effective method for screening archived skin biopsy RNA to determine the prevalence of microbiome components across the SSc population. Improvements in our sample-processing pipeline now allow for simultaneous extraction of DNA, RNA, and miRNA from all patient biopsies. DNA is being used as a template for targeted sequencing of the intergenic transcribed spacer regions (ITS), a region widely regarded as the gold standard for fungal species identification. To date, targeted ITS sequencing libraries have been analyzed from 48 archived samples (39 SSc and 9 controls), which includes both paired lesional and non-lesional skin as well as multiple time points from a single patient (Figure 5). Sequencing outputs are being analyzed by IMSA to identify differences in microbial diversity and species abundance between patients and controls, between lesional and non-lesion skin, as well as how these populations change over time.

Analysis of microbiome component within archived RNA samples was piloted using both nested PCR and NanoString-based methods. Nested PCR-based approaches proved incompatible with RNA, as the ribosomal variable regions proved too large to sequence on the Ion Torrent or similar platforms. Smaller intergenic

regions commonly used for sequenced-based species identification are lost during RNA processing. NanoString-based analyses were capable of distinguishing between fungi to the species level, with no cross-reactivity seen between *R. glutinis* and other closely related species. However, preliminary analyses revealed microbiome components to be below the limit of detection. Targeted amplification prior to NanoString-based analysis will be necessary to overcome these limitations and this option is currently being investigated.

Figure 5. Targeted ITS sequencing of normal and SSc skin biopsies. Below is a preliminary analysis of targeted ITS sequencing which shows a subset of patients have increased *R. glutinis* sequences (red). The most prominent fungal species detected on skin were *Malassezia* spp. (blue), the most common genus of skin commensal fungi.

Proportional Distribution of Fungi On Normal and SSc Skin

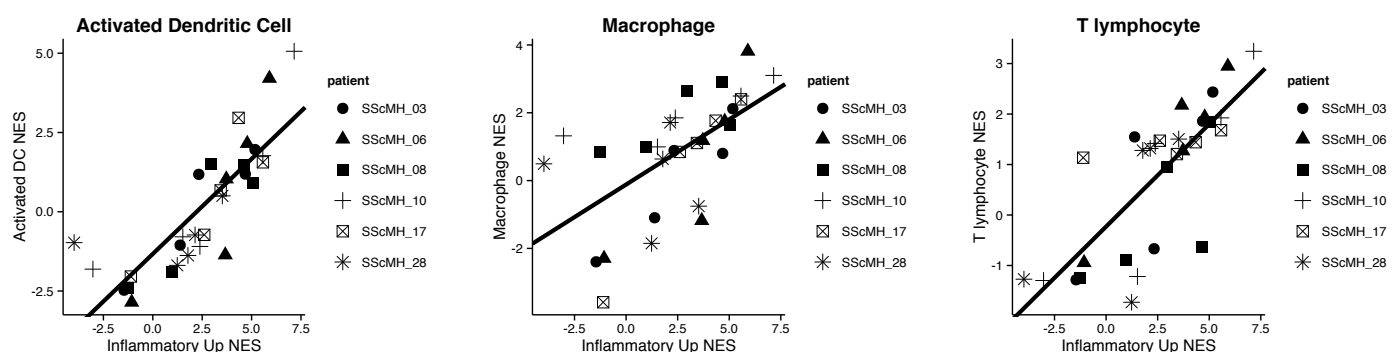


Task 6: Culture microbiome components from the skin of SSc patients. Use of skin biopsies as a method for fungal culture has not been successful, likely due to the use of antiseptics prior to biopsy collection as a means of preventing infection of the biopsy site. In order to streamline patient care, and preserve biopsy composition, skin swabs are being used in lieu of skin scraping as a means of microbial collection prior to biopsy. *Rhodotorula* spp., as well as a variety of other yeast and molds, have been isolated from control patients. Expansion of patient swabbing efforts as a means of fungal detection is ongoing.

Milestone 2: Identify the inflammatory infiltrates in SSc skin and their response to microbiome components

Task 1: Computational prediction of inflammatory cell infiltrates from genomic expression data. We have used single sample Gene Set Enrichment Analysis (ssGSEA) to identify the cellular subsets in SSc skin at different stages of disease. We first benchmarked the ssGSEA method in my laboratory using publicly available gene expression data from pools of cell lines that had a known composition (data not shown). These data demonstrated that ssGSEA accurately predicted cell type enrichment. We then analyzed a set of patients for whom we had whole genome expression data and that had strong expression of the inflammatory signature. We find the inflammatory signature is most strongly correlated with gene expression signatures from activated Dendritic Cells (DCs) and macrophages (MØs) (Figure 6).

Figure 6. Correlation of cell type signatures with a patient's inflammatory signature normalized enrichment score (NES). The inflammatory signature in SSc skin is most highly correlated with activated DCs and MØs.

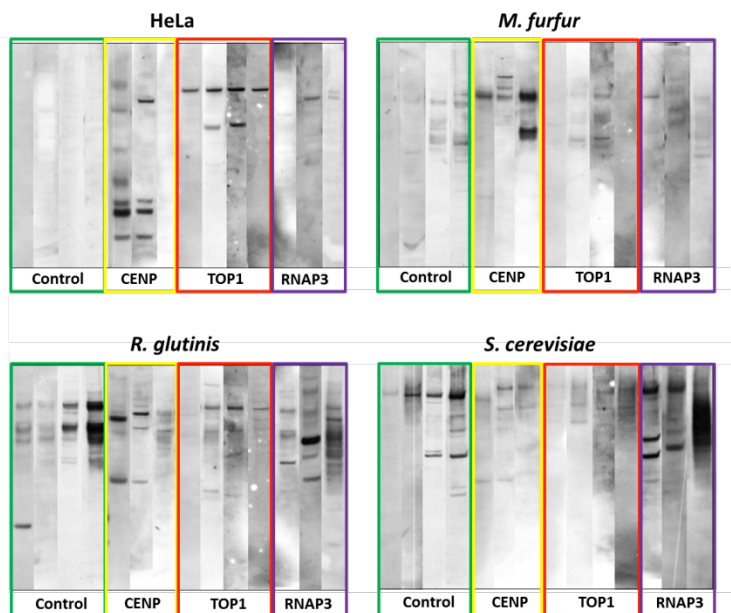


Task 2: Perform immunohistochemistry to validate the computational predictions of task 1 above. We are currently optimizing markers for different cell types in SSc skin. We will use CD163 for macrophages and CD3 for T cells. The precise series of markers is ongoing and being carefully defined.

Milestone 3: Determine if SSc patients have a specific immune response against *R. glutinis* that is different from healthy controls and if this response can drive fibrosis.

Task 1: Test patient sera for cross-reactivity against *R. glutinis* antigens. In these experiments, we set out to test the hypothesis that autoantibody reactivity observed in SSc could recognize the same proteins in fungi, indicating that autoantibodies may have originated in response to fungal infection. Western blots were performed using *R. glutinis*, *Malassezia furfur*, *Saccharomyces cerevisiae*, and HeLa whole cell lysates (to test cross-reactivity with humans), and probed with sera collected from both healthy controls and SSc patients representing the three major autoantibody groups (Controls, CENP, TOP1, and RNAP3). Clear differences in cross-reactivity were evident between patient subsets. SSc patients showed a pattern of cross reactivity against *R. glutinis* lysates that was distinct from that observed in healthy controls. Among clinical autoantibody groups, a band consistent with the presence of TOP1 was seen in 3 of 4 TOP1 patients against *R. glutinis* and HeLa cells (Figure 7) this band was not observed in either *M. furfur* or *S. cerevisiae*, suggesting the possibility of cross-reactivity between *R. glutinis* and human TOP1 (Figure 7). Specific cross reactivity was also observed in CENP and RNAP3 patients; the identity of these proteins is being investigated.

Figure 7. Western blots using SSc and control Sera



Task 2. Identify the cross-reacting proteins by mass spectrometry. Serum-immunoprecipitation of *R. glutinis* and HeLa cell whole cell lysates followed by mass spectrometry was performed to identify immunoreactive proteins associated with *R. glutinis*. We have written a manuscript examining cross-reactivity against human lysates. We are currently having difficulty examining reactivity against *R. glutinis* due to insufficient annotation of the *R. glutinis* genome, limiting our ability to annotate detected spectra.

Serum-immunoprecipitation of *R. glutinis* whole cell lysates followed by mass spectrometry was performed to identify immunoreactive proteins associated with *R. glutinis*. Considerable reactivity was seen for both SSc patients and healthy controls against *R. glutinis* lysates; however, identification of target peptides was not possible due to the absence of a sufficiently well-annotated *R. glutinis* proteome. To overcome this obstacle, a comparable assay using *S. cerevisiae* whole cell lysates was performed, revealing cross-reactivity between major SSc autoantibodies and their fungal homologs.

CONCLUSION:

We have made significant progress on all milestones for the first year of our grant. We have begun sequencing samples, and have performed a preliminary analysis of gene expression changes, as well as a

metagenomic analysis for micro- and mycobiome composition. Preliminary analyses have broadly confirmed our initial results; additional sequencing and analyses are ongoing. Computational analyses have identified activated pDCs and MØs as the key cell types driving the inflammatory signature, a phenotype consistent with the presence of a mycobiome trigger; these results are now being confirmed experimentally. Finally, we have begun to analyze the cross-reactivity of autoantibodies with fungal components. A paper reporting the autoantibody cross-reactivity to human proteins in HeLa cells has been prepared and is submitted.

KEY RESEARCH ACCOMPLISHMENTS Summary

The next reporting period:

September 2015-September 2016

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

The major impact of this project is that we are demonstrating a novel paradigm for the initiation of SSc. This has the potential to dramatically change the way we think about SSc and the role of the innate immune system in driving disease.

What was the impact on other disciplines?

This study impacts areas of genomics, metagenomics, microbiology, innate immunity, and autoimmunity. The methods we demonstrate and develop here will affect all of these fields. In particular, this study begins to develop methods for both systems biology and metagenomic sequencing analyses that can be used in other rare diseases.

What was the impact on technology transfer?

Technical demands associated with this project necessitated the development of a novel method to isolate DNA, RNA, and miRNAs from a single skin biopsy. This method has been submitted as a disclosure to our technology transfer office (TTO).

What was the impact on society beyond science and technology?

Scleroderma is an incurable disease that often has a very poor prognosis. If our metagenomic results are confirmed, this will provide not only a better understanding of the molecular processes driving disease pathogenesis, but also identify alternative strategies, such as anti-fungal treatment, as a possible treatment for SSc.

5. CHANGES/PROBLEMS

None to report.

6. PRODUCTS:

None at this time.

Oral Presentations: (Chronological Order)

Presentations for Michael L. Whitfield, PhD

- | | |
|-------|---|
| 12/15 | <i>"Big Data in the Life Sciences"</i> North Carolina Central University, Durham, NC (<i>Scheduled</i>) |
| 12/15 | <i>"Big Data in the Life Sciences"</i> North Carolina State University, Raleigh, NC (<i>Scheduled</i>) |
| 12/15 | <i>"Big Data in the Life Sciences"</i> University of North Carolina, Chapel Hill, NC (<i>Scheduled</i>) |
| 11/15 | <i>"Multi-tissue genomic networks and systems biology in systemic sclerosis"</i> . Scleroderma Foundation Workshop, ACR Annual meeting, San Francisco CA (<i>Scheduled</i>) |
| 8/15 | <i>"Systems Biology in Systemic Sclerosis."</i> Session Chair and topic introduction. Scleroderma |

Basic Science Workshop, Cambridge UK.

- 6/15 “Defining overlapping pathology between SSc patients and commonly used mouse models of disease” Actelion, Basel Switzerland. (cancelled due to illness)
- 6/15 “Genomic and Proteomic Quantification of the Heterogeneity of SSc: Implications for Pathogenesis and Treatment”. EULAR. Rome, Italy. (cancelled due to illness)
- 6/15 “Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis”. NIH CORT (P50) Advisory Committee meeting. Boston University Medical Center, Boston MA
- 4/15 “Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis”. SScScores (NIH P30) Advisory Committee meeting. Boston University Medical Center, Boston MA.
- 3/15 “A macrophage-associated inflammatory signature is found in all SSc tissues and associated with more severe disease” Scleroderma Research Foundation Workshop on Scleroderma, San Francisco, CA
- 3/15 “Molecular stratification and drug response for SSc clinical trials” Pfizer, Cambridge, MA.
- 2/15 “Enabling Precision Medicine in SSc Clinical Trials” Discussion leader and presenter, NIAMS roundtable discussion on Scleroderma: Advancing Potential Drugs to Patient Care
- 2/15 “Linking autoimmune systemic sclerosis and cancer: disease stratification, co-expression networks and genetic polymorphisms” Cancer Mechanisms Program, Norris Cotton Cancer Center.
- 1/14 “Mechanisms of Systemic Sclerosis (Scleroderma) pathogenesis by systems level genomic analyses” Genomic Medicine Grand Rounds, Geisel School of Medicine.
- 12/14 “Untangling molecular changes in SSc clinical trials: Gene expression subsets, response signatures and pathway changes” ASSET Investigator Meeting. University of Michigan, MI
- 11/14 “Identification of the Microbiome As a Potential Trigger of Systemic Sclerosis By Metagenomic RNA-Sequencing of Skin Biopsies” ACR Basic Research Conference Boston, MA.

Dr. Sarah Arron reports no presentations on this topic in the past year.

Abstracts and Presentations: (Chronological Order)

1. Michael E. Johnson, Zhenghui Li, Michelle T. Dimon, Tammara A. Wood, Robert Lafyatis, Sarah T. Arron, **Michael L. Whitfield**. Identification of the microbiome as a potential trigger of systemic sclerosis by metagenomic RNA-sequencing of skin biopsies. American College of Rheumatology Annual Meeting, 2014
2. Zhenghui Li, Eleni Marmarelis, Kun Qu, Lionel Brooks, Patricia A. Pioli, Howard Y. Chang, Robert Lafyatis, and **Michael L. Whitfield**. RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression. American College of Rheumatology Annual Meeting, 2014

Manuscripts:

The following manuscripts from Dr. Whitfield’s lab have relevance to this proposal. Publication 5 and 6 directly derives in part from work performed to accomplish the aims of this proposal.

1. Arron ST, Dimon MT, Li Z, Johnson ME, Wood T., Feeney L, Angeles JG, Lafyatis R, **Whitfield ML***. High *Rhodotorula* sequences in skin transcriptome of patients with diffuse systemic sclerosis. J. Invest Derm. 2014, Mar 7. doi: 10.1038/jid.2014.127.
2. Johnson ME, Mahoney JM, Marmarelis E, Sargent JR, Wu MR, Spotts K, Hinchcliff M, **Whitfield ML**. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One. 2015 Jan 21;10(1):e0114017. doi: 10.1371/journal.pone.0114017. eCollection 2015.
3. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, Hinchcliff M, **Whitfield ML***. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol. 2015 Jan 8;11(1):e1004005. doi: 10.1371/journal.pcbi.1004005. eCollection 2015 Jan.
4. Taroni JN, Martyanov V, Wood TA, Choe S, Huang CC, Hirano I, Yang GY, Brenner D, Jung B, Carns M, Podlaski S, Chang RW, Varga J, **Whitfield ML**, Hinchcliff M. Genome-wide gene expression analysis of systemic sclerosis esophageal biopsies identifies disease-specific molecular subsets. Arthritis Research & Therapy, (2015) *In Press*
5. Michael E. Johnson, Andrew V. Grassetti, Jaclyn N. Taroni, Shawn M. Lyons, Devin Schweppe, Jessica K. Gordon, Robert F. Spiera, Robert Lafyatis, Paul J. Anderson, Scott A. Gerber, **Michael L. Whitfield**. Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis. Arthritis Research & Therapy, *Submitted*
6. Zhenghui Li, Guoshuai Cai, Michael S. Ball, Kun Qu, Patricia A. Pioli, Howard Chang, Sarah Arron, Robert Lafyatis, and **Michael L. Whitfield**. Functional Characterization of Systemic Sclerosis Transcriptome Identifies a Coding Region Polymorphism more Prevalent in Africans that affects IL6 Production. *In preparation*

The following papers were published by Drs. Whitfield and Lafyatis during the funding period.

7. Long KB, Li Z, Burgwin C, Cho SG, Martyanov V, Sassi-Gaha S, Earl J, Eutsey R, Ahmed A, Ehrlich GD, Artlett CM, **Whitfield ML**, Blankenhorn EP *. The Tsk2/+ mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the *Col3a1* gene. J. Invest Derm. 2014, Oct 20. doi: 10.1038/jid.2014.455.
8. Iwamoto N, Vettori S, Maurer B, Brock M, Jüngel A, Calcagni M, Gay RE, **Whitfield ML**, Distler J.H.W, Gay S, Distler O*. Downregulation of miR-193b in systemic sclerosis regulates the proliferative vasculopathy by urokinase-type plasminogen activator expression. Ann Rheum Dis. 2014 Nov 10. pii: annrheumdis-2014-205326. doi: 10.1136/annrheumdis-2014-205326. [Epub ahead of print]
9. Marangoni RG, Korman B, Wei J, Wood TA, **Whitfield ML**, Scherer PE, Tourtellotte WG and Varga J*. Myofibroblasts in Cutaneous Fibrosis Originate from Intradermal Adipocytes. Arthritis Rheumatol. 2015 Apr;67(4):1062-73. doi: 10.1002/art.38990.
10. Chakravarty EF, Martyanov V, Fiorentino D, Wood TA, Haddon DJ, Jarrell JA, Utz PJ, Genovese MC, **Whitfield ML**, Chung L. A Pilot Randomized Placebo-Controlled study of Abatacept for the Treatment of Diffuse Cutaneous Systemic Sclerosis. Arthritis Research & Therapy, Arthritis Res Ther. 2015 Jun 13;17(1):159.
11. Fresolimumab treatment decreases biomarkers and improves clinical symptoms in systemic sclerosis patients. Rice LM, Padilla CM, McLaughlin SR, Mathes A, Ziemek J, Goummih S, Nakerakanti S, York M, Farina G, **Whitfield ML**, Spiera RF, Christmann RB, Gordon JK, Weinberg J, Simms RW, Lafyatis R. J. Clin. Invest. 2015 Jun 22. pii: 77958. doi: 10.1172/JCI77958

12. Lisa M. Rice, Jessica Ziemack, Eric Stratton, Sarah McLaughlin, Cristina Padilla, Allison Mathes, Romy Christmann, Giuseppina Stifano, Jeff Browning, **Michael L. Whitfield**, Robert Spiera, Jessica Gordon, Robert Simms, Yuqing Zhang, Robert Lafyatis. A Second Generation Pharmacodynamic Biomarker for Diffuse Cutaneous Systemic Sclerosis. *Arthritis and Rheum.* (2015) *In Press*
13. Gordon JK, Martyanov V, Wood TA, Spiera RF, **Whitfield ML**. Nilotinib (Tasigna™) in the Treatment of Early Diffuse Systemic Sclerosis: An Open-Label, Pilot Clinical Trial. *Arthritis Research & Therapy* (2015) *In Press*
14. Brooks L, Lyons SM, Mahoney JM, Welch JD, Liu Z, Marzluff WF, and **Whitfield ML**. A multi-protein occupancy map of the histone mRNP. *RNA* (2015) *In Press*

Degrees obtained that are supported by this award

Dr. Zhenghui Li, who worked on the microbiome and Tsk2/+ projects, will complete his PhD during year 2 of funding. He has received direct support from this grant.

Development of cell lines, tissue or serum repositories

None

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

None

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to <https://ers.amedd.army.mil> for each unique award.

An identical final progress report will be sent from Dr. Arron

9. REFERENCES

1. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML: **Molecular subsets in the gene expression signatures of scleroderma skin.** *PLoS ONE* 2008, **3**:e2696.
2. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML: **Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies.** *J Invest Dermatol* 2012, **132**:1363-1373.
3. Hinchcliff M, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, Tamaki Z, Lee J, Carns M, Podlusk S, et al: **Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis.** *J Invest Dermatol* 2013, **133**:1979-1989.
4. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, Hinchcliff ME, Whitfield ML: **Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms.** *PLoS Comput Biol* 2015, **11**:e1004005.
5. Arron ST, Dimon MT, Li Z, Johnson ME, Wood T, Feeney L, Angeles JG, Lafyatis R, Whitfield ML: **High Rhodotorula Sequences in Skin Transcriptome of Patients with Diffuse Systemic Sclerosis.** *J Invest Dermatol* 2014.

10. APPENDIX

1. Arron ST, Dimon MT, Li Z, Johnson ME, Wood T., Feeney L, Angeles JG, Lafyatis R, **Whitfield ML***. High *Rhodotorula* sequences in skin transcriptome of patients with diffuse systemic sclerosis. J. Invest Derm. 2014, Mar 7. doi: 10.1038/jid.2014.127.
2. Michael E. Johnson, Andrew V. Grassetti, Jaclyn N. Taroni, Shawn M. Lyons, Devin Schweppe, Jessica K. Gordon, Robert F. Spiera, Robert Lafyatis, Paul J. Anderson, Scott A. Gerber, **Michael L. Whitfield**. Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis, *Submitted*

The *Tsk2*/+ Mouse Fibrotic Phenotype Is Due to a Gain-of-Function Mutation in the PIIINP Segment of the *Col3a1* Gene

Kristen B. Long¹, Zhenghui Li², Chelsea M. Burgwin¹, Susanna G. Choe², Viktor Martyanov², Sihem Sassi-Gaha¹, Josh P. Earl³, Rory A. Eutsey³, Azad Ahmed³, Garth D. Ehrlich³, Carol M. Artlett¹, Michael L. Whitfield² and Elizabeth P. Blankenhorn¹

Systemic sclerosis (SSc) is a polygenic, autoimmune disorder of unknown etiology, characterized by the excessive accumulation of extracellular matrix (ECM) proteins, vascular alterations, and autoantibodies. The tight skin (*Tsk2*)/+ mouse model of SSc demonstrates signs similar to SSc including tight skin and excessive deposition of dermal ECM proteins. By linkage analysis, we mapped the *Tsk2* gene mutation to <3 megabases on chromosome 1. We performed both RNA sequencing of skin transcripts and genome capture DNA sequencing of the region spanning this interval in *Tsk2*/+ and wild-type littermates. A missense point mutation in the procollagen III amino terminal propeptide segment (PIIINP) of collagen, type III, alpha 1 (*Col3a1*) was found to be the best candidate for *Tsk2*; hence, both *in vivo* and *in vitro* genetic complementation tests were used to prove that this *Col3a1* mutation is the *Tsk2* gene. All previously documented mutations in the human *Col3a1* gene are associated with the Ehlers–Danlos syndrome, a connective tissue disorder that leads to a defect in type III collagen synthesis. To our knowledge, the *Tsk2* point mutation is the first documented gain-of-function mutation associated with *Col3a1*, which leads instead to fibrosis. This discovery provides insight into the mechanism of skin fibrosis manifested by *Tsk2*/+ mice.

Journal of Investigative Dermatology (2015) 135, 718–727; doi:10.1038/jid.2014.455; published online 20 November 2014

INTRODUCTION

There are multiple animal models of systemic sclerosis (SSc) (Artlett, 2010); yet, none mimics all facets of SSc disease. Of the genetic models, the cause of disease in tight-skin 1 (*Tsk1*/+) mice is known to be a tandem duplication in the fibrillin-1 (*Fbn1*) gene (Siracusa *et al.*, 1996). Other models of SSc have employed mice with individual gene deficiencies or overexpression including Fos-related antigen-2 (*Fra2*; Maurer *et al.*, 2009), endothelin-1 (*Edn1*; Hochoer *et al.*, 2000; Richard *et al.*, 2008), and Friend leukemia integration 1 transcription factor (*Fli1*; Asano *et al.*, 2010), which have proven useful for understanding the contribution of these proteins to the

vasculopathy and/or lung fibrosis seen in SSc. Nongenetic models of SSc include the bleomycin-induced scleroderma model (Yamamoto *et al.*, 1999), which has been used to study many of the initiating events involved in fibrosis.

The *Tsk2*/+ mouse was first described in 1986, when an offspring of a 101/H mouse exposed to the mutagenic agent ethylnitrosourea was noted to have tight skin in the interscapular region (Peters and Ball, 1986). The mutagenized gene causing SSc-like signs in *Tsk2*/+ mice was reported to be located on chromosome 1 between 42.5 and 52.5 megabases (Mb; Christner *et al.*, 1996); however, the genetic defect was never identified. Similar to *Tsk1*, *Tsk2* SSc-like traits are highly penetrant in *Tsk2*/+ heterozygotes and it is homozygous embryonic lethal. *Tsk2*/+ mice have many features of human disease including tight skin, dysregulated dermal extracellular matrix (ECM) deposition, and evidence of an autoimmune response (Christner *et al.*, 1995; Gentiletti *et al.*, 2005).

Herein, we report the positional cloning and identity of the *Tsk2* gene. We have discovered that *Tsk2*/+ mice carry a deleterious gain-of-function missense mutation in *Col3a1* (collagen, type III, alpha 1), which exchanges a cysteine for serine in the N-terminal propeptide, procollagen III amino terminal propeptide segment (PIIINP). The *Tsk2*/+ mouse affords a unique opportunity to examine the pathways leading to the multiple clinical parameters of fibrotic disease from birth onward.

¹Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA; ²Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA and ³Center for Genomic Sciences, Pittsburgh, Pennsylvania, USA

Correspondence: Elizabeth P. Blankenhorn, Department of Microbiology and Immunology, Drexel University College of Medicine, 2900 Queen Lane, Philadelphia, Pennsylvania 19129, USA.

E-mail: Elizabeth.Blankenhorn@drexelmed.edu

Abbreviations: B6, C57Bl/6; *Col3a1*, collagen, type III, alpha 1; ECM, extracellular matrix; KO, knockout; Mb, megabase; PIIINP, procollagen III amino terminal propeptide segment; SNP, single-nucleotide polymorphism; SSc, systemic sclerosis; *Tsk*, tight skin; UTR, untranslated region; WT, wild type

Received 13 June 2014; revised 9 September 2014; accepted 22 September 2014; accepted article preview online 20 October 2014; published online 20 November 2014

RESULTS

Linkage and sequencing studies reveal a SNP mutation in *Col3a1*

Identification of the *Tsk2* gene was initiated with further mapping of the *Tsk2* interval by genotyping backcross progeny of *Tsk2*/+ mice bred to C57Bl/6 (B6) mice. Littermate mice were genotyped for informative microsatellites (*D1Mit233*, *D1Mit235*, a microsatellite in *Gls*, and *D1Mit18*) and single-nucleotide polymorphism (SNP) genotyping assays used for additional markers. Multiple recombinants were recovered that mapped the interval to between 42.53 and 52.22 Mb on chromosome 1. Recombinants were bred and then backcrossed to a consomic B6.chr 1-A/J mouse to fine map the region by SNP typing, as A/J mice bear many known SNPs compared with B6 mice. Additional recombinants were recovered and new SNPs from the sequencing projects (see below) were used to narrow the *Tsk2* interval to between 44.67 and 46.27 Mb (Figure 1a), representing a >3-fold reduction in the size of the interval bearing 101/H genomic DNA and *Tsk2*. There are six known genes in this interval (Figure 1b).

To identify the mutation underlying *Tsk2*, we employed both RNA sequencing (RNA-Seq) and genome capture sequencing of the reduced genomic interval. Sequence reads were aligned to the MM9 reference genome (B6) and analyzed for polymorphisms in the *Tsk2* interval. There were 265 SNPs found in both wild type (WT) and *Tsk2*/+ littermates that represent differences between the reference B6 genome and the 101/H background; these were excluded from further study. Thirteen SNPs were found in all four *Tsk2*/+ mice analyzed; 10 of these SNPs were also found to be in liver RNA from 101/H strain or in other non-fibrotic mouse strains (<http://phenome.jax.org/>) and were also ruled out as candidates for *Tsk2* (Table 1). The remaining three SNPs were heterozygous and confirmed to be only in *Tsk2*/+ mice. One of these, in a

Gulp1 intron, proved useful as an additional marker that resides outside the supported linkage interval for *Tsk2*/+ on the proximal end in an informative recombinant mouse (Figure 1a). A second SNP was also found in an intron of *Gulp1*. The RNA-Seq data did not identify any splicing defects in *Gulp1* mRNA in the *Tsk2*/+ mice (Supplementary Figure S1 online), indicating that this SNP does not change *Gulp1* mRNA splicing, and its gene expression in the skin is unchanged (Figure 2). Thus, the intronic SNP in *Gulp1* is unlikely to have a role in the tight skin phenotype. The remaining mutation was in *Col3a1* that results in a T-to-A transversion at Chr1:45,378,353, causing a Cys→Ser amino acid change in the PIINP, a natural cleavage product of COL3A1. The mutant protein is designated COL3A1^{*Tsk2*} (C33S).

We calculated the reads per kilobase per million mapped reads for each gene and found that of the genes in the reduced genomic interval, *Col3a1* shows the highest absolute expression level with all other genes showing negligible expression levels. RNA-Seq results indicate that there is a trend toward higher *Col3a1* mRNA abundance in 4-week-old *Tsk2*/+ skin samples compared with WT littermates (Figure 2a and b). The *Col3a1*^{*Tsk2*} (C33S) mutation is unlikely to change the expression levels of the *Col3a1* mRNA directly but will result in a mutated protein that is deposited in the ECM along with the WT protein in mixed heterotrimers, and could result in activation of pathways that impinge on *Col3a1*, such as transforming growth factor- β (Sargent, *et al.*, submitted). Because *Tsk2*/+ (affected) mice are heterozygous, the *Col3a1*^{*Tsk2*} (C33S) mutation should account for 50% of the reads assuming equal expression from each allele. We calculated the read count from the RNA-Seq data for the reference and alternate alleles for *Col3a1* at Chr1:45,378,353. In WT mice, we find that all reads (492 total) contain the

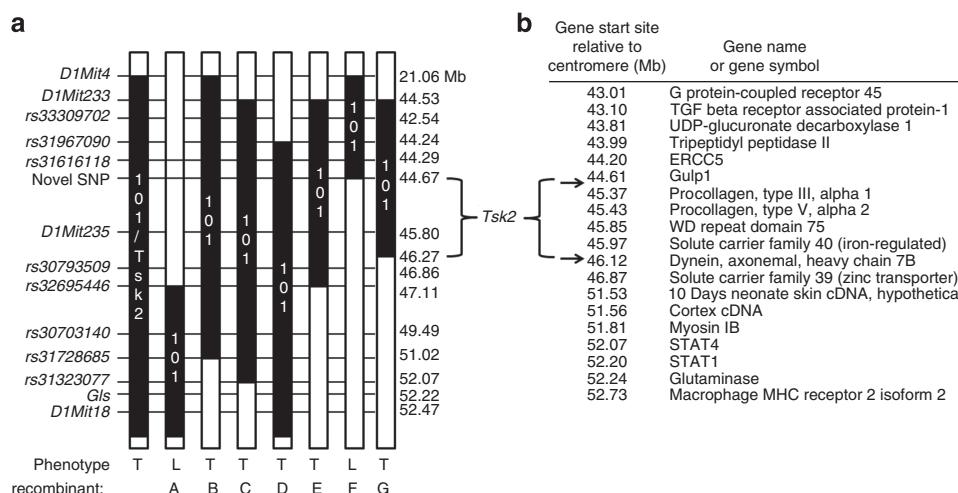


Figure 1. *Tsk2* lies between and not including 44.67–46.27 Mb on chromosome 1. (a) The *Tsk2* interval was narrowed by genotyping backcrossed mice on the B6 and B6.chr 1-A/J backgrounds. Black bars (101/H) depict the original parental strain, bearing *Tsk2*. White bars depict the B6 genome. Recombinants A–G bear additional recombination sites. The phenotypes are tight (T—*Tsk2*/+) or loose (L—WT). (b) With the use of additional markers (arrows, see text), the current interval comprises *Col3a1*, *Col5a2*, *Wdr75*, *Slc40a1*, part of *Gulp1*, and part of *Dnahe7b*; the five latter genes do not have coding region mutations. The elements of the *Gulp1* gene above 44.67 Mb are excluded by the recombination in mouse F, and *Dnahe7b* below 46.27 is excluded by mouse G. B6, C57Bl/6; *Col3a1*, collagen, type III, alpha 1; Mb, megabases; SNP, single-nucleotide polymorphism; Tsk, tight skin; WT, wild type.

Table 1. Nucleotide changes between *Tsk2*⁺ mice and 101/H or B6 mice

Nucleotide position on Chr 1 (MM9)	Genotype of <i>Tsk2</i> ⁺	Genotype of B6	Genotype of 101/H	Present in other strains?	Potential candidate for <i>Tsk2</i> ?	Gene or mRNA containing substitution
<i>SNP found by RNA-Seq</i>						
44,675,490	A	T	T	No	No, outside interval	<i>Gulp1</i> intron
44,833,682*	C	T	T	No	Yes	<i>Gulp1</i> Intron
45,378,353*	A	T	T	No	Yes	<i>Col3a1</i> exon (C33S)
45,432,389	C	G	ND	Yes	No	<i>Col5a2</i> 3'UTR
45,441,243	C	A	C	No	No, in 101/H	<i>Col5a2</i> intron
45,860,529	G	A	G	Yes	No	<i>Wdr75</i> intron
45,874,790	T	C	T	Yes	No	<i>Wdr75</i> intron
45,875,728	C	T	C	Yes	No	<i>Wdr75</i> exon
45,880,257	CG	AC	CG	No	No, in 101/H	<i>Wdr75</i> exon
46,872,610	T	G	ND	Yes	No	<i>Slc39a10</i> intron
46,874,711	C	T	C	Yes	No	<i>Slc39a10</i> intron
46,939,340	T	C	T	Yes	No	BC040767 intron
46,939,624	A	G	ND	Yes	No	BC040767 intron
<i>SNP found by Genome Capture Sequencing (454)</i>						
44,833,682*	C	T	T	No	Yes	<i>Gulp1</i> intron
45,378,353*	A	T	T	No	YES	<i>Col3a1</i> exon (C33S)
45,465,923	A	T	T	No	YES	<i>Col5a2</i> intron
46,124,856	A	G	A	Yes	No	<i>Dnahc76</i> intron
46,124,857	A	C	T	Yes	No	<i>Dnahc76</i> intron
46,268,651	C	T	T	No	No, outside interval	<i>Dnahc76</i> intron

Abbreviations: B6, C57Bl/6; Chr, chromosome; Col3a1, collagen, type III, alpha 1; ND, not determined; RNA-Seq, RNA sequencing; SNP, single-nucleotide polymorphism; Tsk, tight skin.

All single-copy nucleotide changes found by RNA-Seq or 454 sequencing were checked for their presence in other non-fibrotic strains (<http://phenome.jax.org/>) or individually verified by a phototyping assay (Bunce *et al.*, 1995) and/or resequencing to confirm the single-nucleotide change. SNPs that were ruled out by one of these assays are considered not to be potential candidates for *Tsk2*. When known, genotypes shown for 101/H are from RNA-Seq, 454 sequencing, or phototyping. *, Seen in both assays.

reference T allele, whereas in *Tsk2*⁺, we find that 48% of reads (273/564 total reads) contain the WT (T) allele and 52% (291/564 total reads) contain the *Col3a1*^{*Tsk2*} (C33S) allele (T > A; Figure 2c). As a comparison, we show that the intronic *Gulp1* SNP at Chr1:44,833,682 has significantly lower read coverage consistent with its intronic location (11-fold coverage in *Tsk2*⁺ and 2-fold coverage in WT). The intronic *Gulp1* SNP also shows a distribution of reads consistent with heterozygosity in *Tsk2*⁺ and with homozygosity in WT (Figure 2d). These findings show that the *Col3a1*^{*Tsk2*} (C33S) locus is heterozygous as expected for the *Tsk2* mutation in these animals, and expression occurs equally from each of the alleles.

Because RNA-Seq only captures variation in the transcribed regions of the genome, and thus might miss an important genomic feature that is unique to *Tsk2*, we sequenced captured genomic DNA samples corresponding to the minimal linkage region from B6.*Tsk2*⁺ heterozygotes and 101/H homozygous parental strain mice. Multiple DNA differences between the *Tsk2*⁺ mouse and its parental 101/H strain were detected. A majority of the differences observed were

accounted for by non-chromosome 1 repetitive DNA sequences such as LINE, SINE, and retroviral elements contained within the *Tsk2* interval on chromosome 1. After filtering repetitive elements from the comparison, there were six single-copy DNA sequence differences, of which three were confirmed to be *Tsk2*⁺ specific (Table 1). Among these, there is a SNP that proved useful in demarcating the distal end of the *Tsk2* linkage interval (Chr1:46,268,651; Table 1 and Figure 1), as it was outside the linkage interval. This allowed us to eliminate the only other gene expressed at an appreciable level in the broader interval, *Slc39a10*. In addition, the GULP1 intronic SNP was confirmed and another SNP in an intron of *Col5a2* was observed. Both these latter SNPs are deemed unrelated to the phenotype, again because of their low overall expression and the lack of any influence on splicing or expression in the RNA-Seq results (Figure 2a and b; Supplementary Figure S1 online). Most importantly, however, the heterozygous T-to-A transversion in *Col3A1* at Chr1:45,378,353 was observed in the genomic sequence comparison and was identical to the mutation identified by RNA-Seq. There were no additional variants that could be

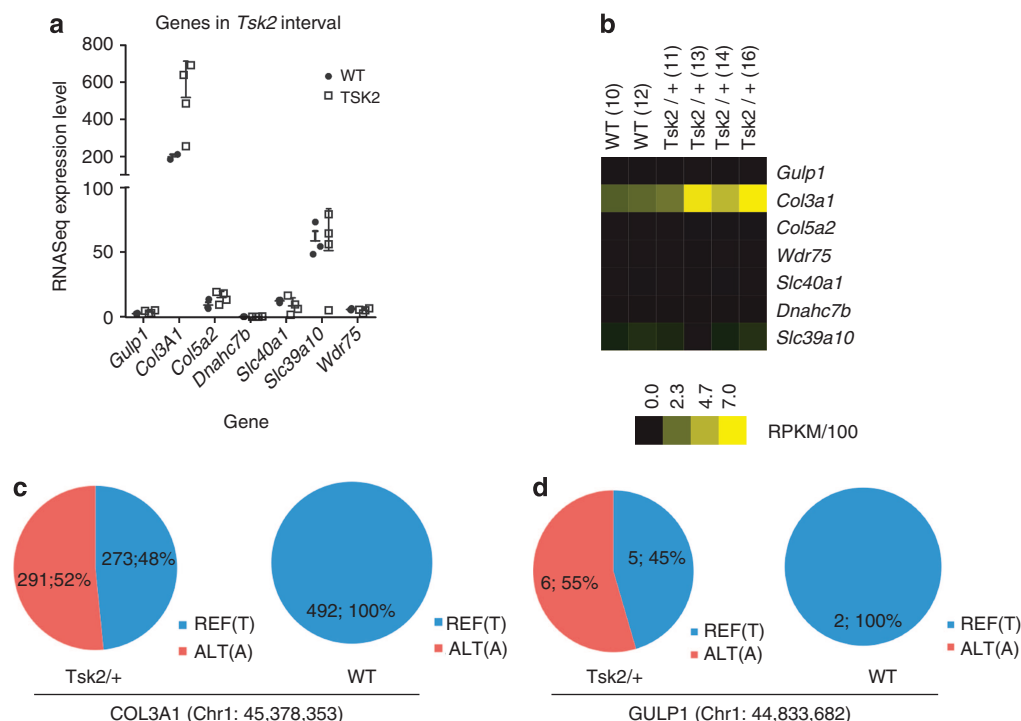


Figure 2. *Col3a1* is the only interval gene expressed at high levels in the skin of *Tsk2/+* mice. (a) This graph shows gene expression for the seven *Tsk2* interval genes, as determined from the RNA-Seq abundance results. (b) Heat map for seven *Tsk2* interval genes detected as transcripts in RNA-Seq. (c, d) Distribution of nucleotide calls in heterozygous *Tsk2/+* and homozygous WT mice for *Col3a1* and *Gulp1*. *Col3a1*, collagen, type III, alpha 1; RNA-Seq, RNA sequencing; *Tsk*, tight skin; WT, wild type.

validated on the *Tsk2* chromosome within ~535,000 nucleotides proximal to the transcription start site of *Col3a1* gene or closer than 59,732 nucleotides distal of the end of the *Col3a1* 3' untranslated region (UTR). Selective resequencing of the 3'UTR likewise revealed no differences between *Tsk2* and 101/H (not shown). Thus, this non-synonymous coding mutation is most likely to be *Tsk2* by genomic assessment, as well as by RNA-Seq.

Mice bearing *Col3a1*^{*Tsk2*} and *Col3a1*^{KO} are not viable

To prove that *Tsk2* is a single-nucleotide change in the *Col3A1* coding region required a separate genetic test. Both *Tsk2/Tsk2* (Peters and Ball, 1986) and *Col3a1*-knockout (KO) (Liu et al., 1997) homozygotes exhibit embryonic lethality, which is also seen in our mouse colony (Table 2). We therefore designed a genetic complementation test to determine whether *Col3a1*^{*Tsk2*} (from *Tsk2* mice) could complement and rescue the null allele for *Col3a1*. Conversely, this same cross would determine whether any other gene in the *Col3a1*-homozygous KO could serve to complement the *Tsk2* mutation.

Tsk2/+ × *Col3a1*^{-/-} mice were bred together, and 37 progeny mice (Table 2) were genotyped. If *Col3a1*^{*Tsk2*} (C33S) can complement the *Col3a1*-KO, then we would expect to find 9 or 10 *Col3a1*^{*Tsk2*}/*Col3a1*-KO compound heterozygotes. In fact, no viable compound heterozygotes were born (Table 2, Supplementary Figure S2 online). The hybrid bearing *Tsk2/Col3a1*-null chromosomes was not viable because the *Tsk2* gene on the *Tsk2*-bearing chromosome

cannot “complement” (rescue) the loss of the *Col3a1* gene on the *Col3a1*-KO chromosome. It bears only the allele of *Col3a1*^{*Tsk2*} at the *Col3a1* locus, which is insufficient to provide a functional COL3A1 protein that is missing in the *Col3a1*-KO. The *Col3a1*-null chromosome likewise cannot complement the *Tsk2* mutation: the remaining genes on the *Col3a1*-KO chromosome cannot prevent the death of (cannot “complement”) mice bearing the *Tsk2* chromosome, whereas hybrids carrying *Tsk2/Col3a1*-WT alleles are alive but fibrotic. In fact, having the *Tsk2* mutation is more damaging than not expressing COL3A1 at all, because, although a few *Col3a1*-KO homozygotes make it to birth, *Tsk2/Tsk2* homozygotes (and *Tsk2/Col3a1*-KO) never do, and, whereas *Col3a1/Tsk2* mice are viable but small in stature and fibrotic, *Col3a1*^{-/+} heterozygotes are normal. Therefore, the mutation in *Tsk2/+* mice lies within *Col3a1* and, when homozygous, is substantially more deleterious compared with a complete genetic deficiency of COL3A1.

Col3a1^{*Tsk2*} induces increased COL1A1 and ECM production *in vitro*

Because the compound heterozygous animals do not survive to accumulate fibrotic levels of ECM, a direct *in vivo* test for fibrosis is impossible; thus, we performed an “*in vitro* complementation” test, wherein we transfected mutant or WT *Col3a1* complementary DNA (cDNA) into *Col3a1*-KO fibroblasts, harvested from a *Col3a1*-KO/KO homozygote at birth. Using the production of COL1A1 as a measure of fibrosis

Table 2. Progeny born from <i>Col3a1</i> -deficient, <i>Col3a1</i> -sufficient, and <i>Tsk2</i> / + mice				
(A) Parents	Genotype and phenotype of progeny			
	<i>Tsk2</i> / + (tight skin)	WT/WT (normal skin)	<i>Tsk2</i> / <i>Tsk2</i> (lethal)	
<i>Tsk2</i> / + × <i>Tsk2</i> / +	22	21	0	
	<i>Col3a1</i> ⁺ / <i>Col3a1</i> [−] (normal skin)	<i>Col3a1</i> ⁺ / <i>Col3a1</i> ⁺ (normal skin)	<i>Col3a1</i> [−] / <i>Col3a1</i> [−] (moribund)	
<i>Col3a1</i> ^{−/+} × <i>Col3a1</i> ^{−/+}	16	13	3	
(B) Parents	Genotype and phenotype of progeny			
	WT/ <i>Col3a1</i> ⁺ (normal skin)	<i>Tsk2</i> / <i>Col3a1</i> ⁺ (tight skin)	WT/ <i>Col3a1</i> [−] (normal skin)	<i>Tsk2</i> / <i>Col3a1</i> [−]
<i>Tsk2</i> / + × <i>Col3a1</i> ^{−/+}	12	10	15	0
Abbreviations: <i>Col3a1</i> , collagen, type III, alpha 1; SNP, single-nucleotide polymorphism; <i>Tsk</i> , tight skin; WT, wild type. All progenies were assessed for chromosome 1 markers (SNPs and microsatellites) that characterize the origin of the tested allele (<i>Tsk2</i> or <i>Col3a1</i>). (A, top) shows the number of mice born of each genotype and phenotype from <i>Tsk2</i> / + × <i>Tsk2</i> / + or <i>Col3a1</i> ^{−/+} × <i>Col3a1</i> ^{−/+} parents. (B, bottom) shows the number of mice born of each genotype and phenotype from <i>Tsk2</i> / + × <i>Col3a1</i> ^{−/+} parents; note: there are no compound heterozygotes (<i>Tsk2</i> / <i>Col3a1</i> [−]) born from this mating.				

(shown to be expressed at high levels in *Tsk2*/+ skin and used as a marker of fibrosis (Barisic-Dujmovic *et al.*, 2008; Christner *et al.*, 1998)), we assessed both protein and mRNA levels in fibroblasts that received DNA from a plasmid containing a single allele of a single *Col3a1* gene. In three independent experiments, COL1A1 protein was significantly elevated after 48 hours of transfection with *Col3a1*^{Tsk2} relative to transfection with *Col3a1*^{WT} (Figure 3a); mRNA for *Col1a1* was likewise increased in cells transfected with mutant *Col3a1*^{Tsk2} cDNA (Figure 3b). Transfection efficiencies were equal in each of the experiments (Figure 3c).

Given the observation that the production of a major indicator of fibrosis, COL1A1, is increased by the transfection of the *Col3a1*^{Tsk2} gene, we assessed the impact of the mutant gene genome-wide. RNA from the *Col3a1*^{Tsk2} and *Col3a1*^{WT} transfected *Col3a1*-KO fibroblasts and from 4-week-old *Tsk2*/+ and WT littermate skin was analyzed by DNA microarray. Differentially expressed pathways between the two transfections were determined by Gene Set Enrichment Analysis (GSEA). Transfection of *Col3a1*^{Tsk2} results in significant enrichment of genes associated with fibrotic Gene Ontology terms including basement membrane, extracellular matrix, integrin binding, and transmembrane receptor protein kinase activity (Figure 3d; GSEA FDR < 5%). The biological processes observed in the skin of four 4-week-old female *Tsk2*/+ mice relative to WT littermates also show increases in genes associated with Gene Ontology terms extracellular matrix, integrin binding, and basal lamina (ZL, CB, KBL, CMA, EPB, MLW, manuscript in preparation). The genes that significantly contributed to the GSEA pathway enrichment in the transfected fibroblasts were extracted from microarray data of the transfections, as well as from female *Tsk2*/+ and WT skin at 4 weeks of age (Figure 3e and f), and were elevated both in the fibroblasts transfected with *Col3a1*^{Tsk2} and in *Tsk2*/+ mouse skin. These include those genes typically associated with fibrosis including *CTGF*, *THY1*, *FBN1*, the collagens, laminins, *TGFβ1*, *TGFβR1*, *ADAMTS* family genes, and *MMP11*. In addition, there was upregulation in *Col3a1*^{Tsk2}-transfected fibroblasts and *Tsk2*/+ skin RNA of the vascular endothelial

growth factor receptors *FLT1* and *FLT4*, as well as genes associated with platelet-derived growth factor signaling (PDGFRB and PDGFRL; Figure 3f). These data indicate that expression of the *Col3a1*^{Tsk2} gene alone can induce a substantial fibrotic gene expression program.

Taken together, this means that *Col3a1* and *Tsk2* are almost certainly one and the same gene. *Col3a1*^{Tsk2} (C33S) is therefore deemed a deleterious gain-of-function allele of *Col3a1*, and the *Col3a1*-KO is a classical loss-of-function allele. Mice thus need at least one copy of a functional, normal *Col3a1* gene.

***Tsk2*/+ mice have increased dermal COL3A1 protein accumulation**

The behavior of *Col3a1* in *Tsk2*/+ mice could reveal the mechanism by which this mutation causes very substantial ECM fibrosis and very tight skin. We measured the level of COL3A1 protein by histological examinations of *Tsk2*/+ and WT littermate skin. Reticular fibers are composed primarily of COL3A1 and are a structural element in the skin, found in the panniculus carnosus and in the dermis. COL3A1 expression in the skin from 2-week-old mice is high and declines after birth in WT littermates but does not decline in the *Tsk2*/+ mice (Figure 4a). As *Tsk2*/+ mice age, the reticular fibers thicken and become more pronounced compared with their WT littermates reflecting the accumulation of COL3A1. This finding was confirmed in the skin from 4-week-old mice by western blots, which revealed that there is significantly more COL3A1 in the skin of *Tsk2*/+ mice compared with age- and sex-matched WT littermates (Figure 4b and c). We propose that the excess COL3A1 protein we observe by several measures in *Tsk2*/+ mice is due to a trend for excess production of *Col3a1* mRNA (Figure 2a) rather than reduced degradation of the Col3 protein. Because the PIIINP fragment is removed from the majority of Col3 molecules before natural Col3 turnover degradation takes place in the tissue, mature COL3A1 from *Tsk2* is identical to mature COL3A1 from WT mice, and its natural degradation is unlikely to be affected by any changes in PIIINP. These data show that there is an overall

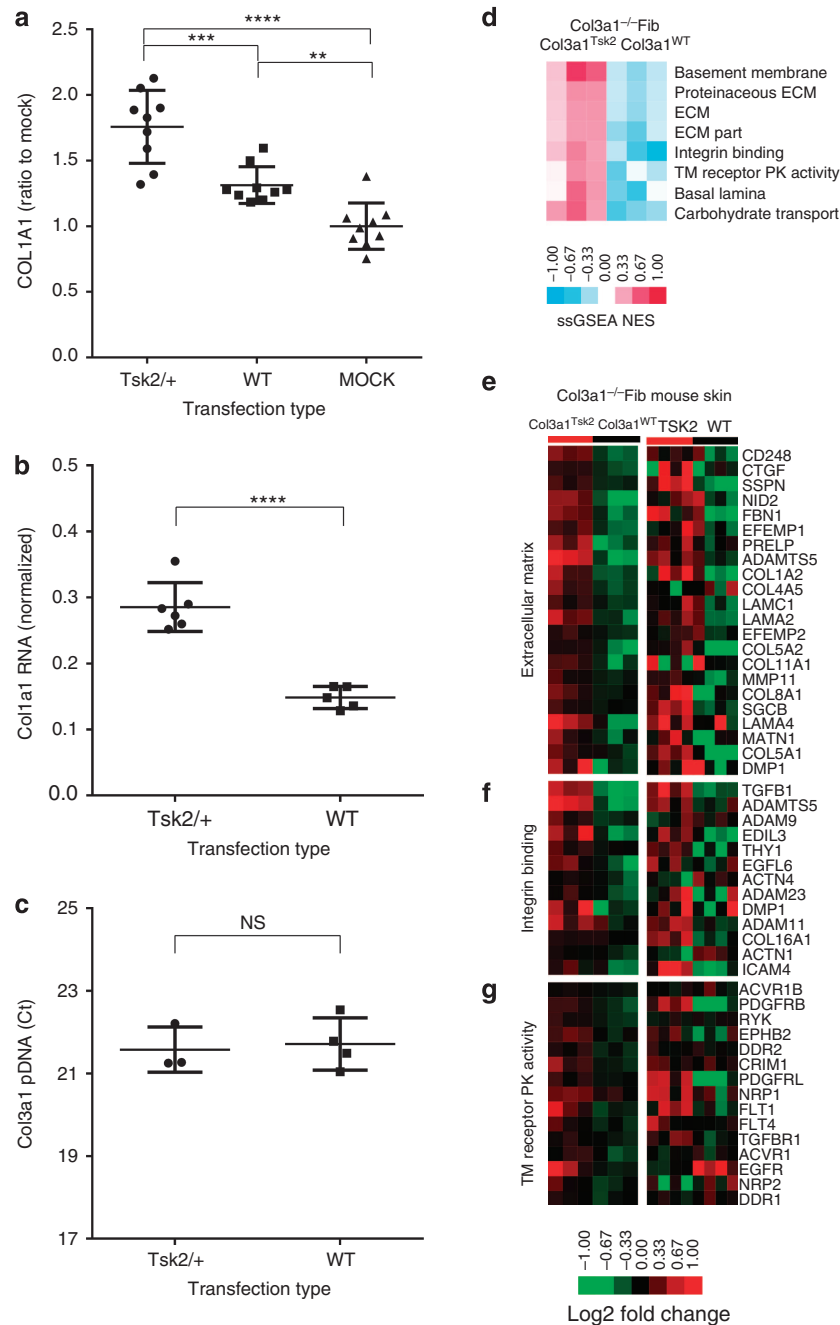


Figure 3. Mouse *Col3a1*-KO fibroblasts transfected with mutant *Col3a1*^{Tsk2} express a more fibrotic protein profile compared with *Col3a1*^{WT} transfectants. (a) Culture supernatants assayed by western blot for COL1A1. *Col3a1*^{Tsk2} transfectants produced 34% more COL1A1 compared with *Col3a1*^{WT} ($P < 0.001$) or mock transfectants ($P < 0.0001$). (b) *Col1a1* mRNA is more highly expressed in *Col3a1*-KO fibroblasts transfected with *Col3a1*^{Tsk2} than with *Col3a1*^{WT} ($P < 0.0001$). (c) There was no significant difference in efficiency of plasmid transfection between *Col3a1*^{Tsk2} and *Col3a1*^{WT}. (d) *Col3a1*^{-/-} fibroblasts transfected with *Col3a1*^{Tsk2} show a significant increase in Gene Ontology terms associated with fibrosis. (e) Expression of the genes that contributed most to the ECM enrichment results in *Col3a1*^{Tsk2} versus *Col3a1*^{WT}-transfected mouse fibroblasts or in 4-week-old female *Tsk2/+* versus WT mice. (f) Expression of genes that contributed to integrin binding term. (g) Expression of genes that contributed to transmembrane receptor protein kinase activity term. *Col3a1*, collagen, type III, alpha 1; ECM, extracellular matrix; KO, knockout; NS, not significant; pDNA, plasmid DNA; Tsk, tight skin; WT, wild type.

increased accumulation of mature COL3A1 protein in the *Tsk2/+* mice; in addition, at least half of the type III procollagen and PIIINP trimers produced likely contain one or more strands bearing the *Tsk2* (C33S) mutation.

DISCUSSION

Sequencing of both expressed RNAs and the genomic region in the *Tsk2/+* interval, coupled with the genetic complementation study, prove that *Tsk2/+* mice harbor a deleterious

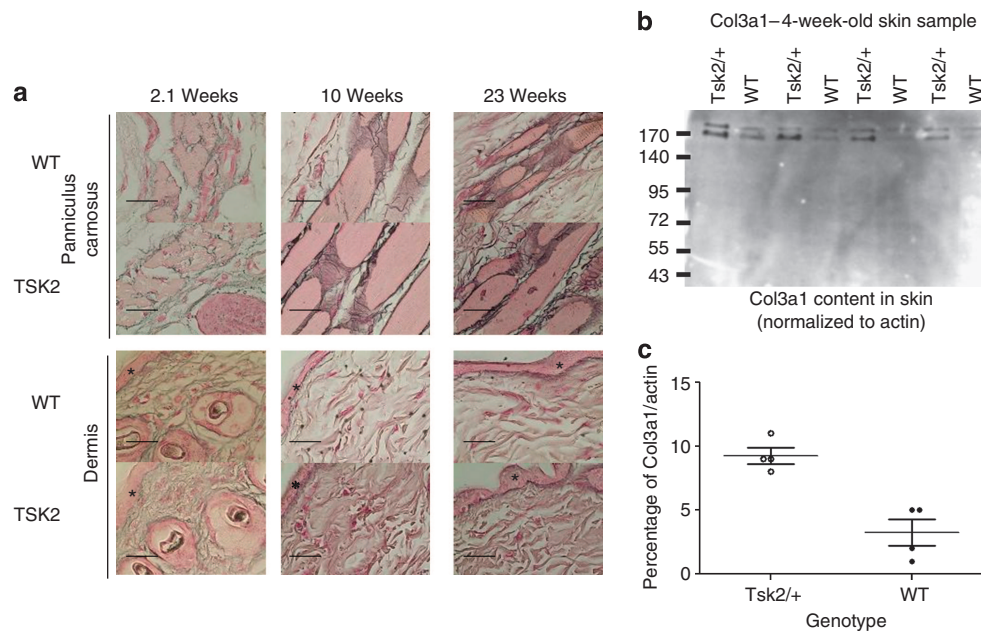


Figure 4. *Tsk2*⁺ mice have increased reticular fiber accumulation and COL3A1 in the skin compared with WT littermates. (a) Reticular fiber staining was performed on mice of the indicated ages (2–23 weeks). Stars mark the location of the epidermis. COL3A1 fibers (black staining) are much thicker and more abundant at each life stage in *Tsk2*⁺ than in WT. Fibers were found to be especially pronounced in the panniculus carnosus region of the tissue; increased staining of COL3A1 in the dermis was also noted. The dermal reticular fibers are composed entirely of COL3A1 protein, as this protein is receptive to silver impregnation, and they are increased in *Tsk2*⁺ mice. All images were taken at 200 \times magnification. Bar size = 100 μ m. (b, c) Skin lysates were analyzed for COL3A1 content (both bands) relative to β -actin (not shown) by western blot analysis. *Tsk2*⁺ mouse skin has significantly more COL3A1 protein than WT mouse skin ($P = 0.0025$, ANOVA). ANOVA, analysis of variance; Col3a1, collagen, type III, alpha 1; Tsk, tight skin; WT, wild type.

coding mutation in *Col3a1*, leading to an amino acid change (C33S) in the N-terminal region of the protein (PIIINP). This point mutation is consistent with those expected from ethylnitrosourea-induced mutagenesis, which generates random single-base-pair point mutations by direct alkylation of nucleic acids. The most common mutations are AT-to-TA and AT-to-GC changes (Noveroske *et al.*, 2000; Cordes, 2005); all three *Tsk2*-specific mutations identified here were T-to-A or T-to-C mutations. The *Tsk2*⁺ allele is expressed in a 1:1 ratio with the WT by RNA-Seq indicating equal transcription and making a duplication event unlikely.

Effects of the *Tsk2* mutation include the following: (1) accumulation of COL3A1 protein *in vivo* over time; (2) induction and accumulation of COL1A1 protein *in vivo* and in *in vitro* expression models; (3) a more lethal phenotype compared with the homozygous genetic loss of *Col3a1*; and (4) a more lethal compound heterozygous phenotype compared with that of the homozygous gene KO. The latter two characteristics indicate that COL3A1^{Tsk2} (C33S) has a dominant prenatal lethal effect, although our *in vitro* complementation results suggest that the presence of COL3A1-C33S (or its mRNA) is not lethal to skin fibroblasts *per se*. A major function of the *Col3a1* gene is promoting blood vessel development (Liu *et al.*, 1997), which likely led to the lethality observed in the complementation experiment. In the *Col3a1*-KO, a few mice are born with the homozygous deficiency, and these mice die of rupture of the major blood vessels (Liu *et al.*, 1997). The possibility that *Col3a1*^{Tsk2} mutation could directly induce a deleterious vascular phenotype in *Tsk2*⁺ mice is

intriguing; it is notable that genes encoding vascular features (*Flt1* and *Flt4*, genes for vascular endothelial growth factor receptors) are significantly upregulated in both *Col3a1*^{Tsk2}-transfected skin fibroblasts and in *Tsk2*⁺ skin relative to WT (Figure 3g). It is possible that a complete *Col3a1* deficiency could be compensated by other collagens, but the *Col3a1*^{Tsk2} mutation is a deleterious gain-of-function, and the deposition of COL3A1-C33S may actively prevent other more benign collagen alternatives from functioning in the vasculature. Thus, our theory is that two doses of a damaging protein are worse than no expression of a normal one.

To our knowledge, this is the first mutation in *Col3a1* that results in a gain-of-function phenotype instead of Ehlers–Danlos-like syndromes that are due to loss-of-function or antimorphic collagen-poor phenotypes. Ehlers–Danlos is a group of connective tissue disorders characterized by highly elastic, fragile but not fibrotic skin due to a defect in collagen synthesis (Nishiyama *et al.*, 2001). In addition, these patients have a significant risk for aneurism. The Ehlers–Danlos syndrome has been associated with 337 mutations in COL3A1 (<http://www.le.ac.uk/ge/collagen/>), as well as mutations on COL1A1 and COL5A2. These mutations result in amino acid substitutions in the C terminus of the protein, RNA splicing alterations, deletions, or null alleles. Interestingly, in the Ehlers–Danlos syndrome type IV (a very different disease than that observed in *Tsk2*⁺ mice), studies have shown that patients bearing a mutated COL3A1 (compared with a null COL3A1) develop more severe disease and succumb to disease prematurely, whereas those with null COL3A1 were

able to live a relatively normal life with limited disease (Leistritz *et al.*, 2011). Currently, all reported COL3A1 mutations result in decreased collagen protein secretion leading to variably thinner skin and defects in the vasculature that are observed in these patients. In contrast to the mutations observed in Ehlers–Danlos, the Tsk2/+ mouse mutation results in thickened skin with no apparent evidence of aneurism. The mutation reported here occurs in the N-terminal PIIINP fragment of the protein, rather than the C-terminal region associated with Ehlers–Danlos.

The PIIINP molecule is a homotrimer with a molecular weight of ~42,000 daltons and comprises three domains: a cysteine-rich globular domain (Col 1) containing 79 amino acids with five intrachain disulfide bonds, a triple-helical domain (Col 3) with 12 amino acids and three interchain disulfide bonds, and a non-collagenous domain (Col 2) comprising 39 amino acids ending with the N-telopeptide that forms a triple helical structure (Bruckner *et al.*, 1978). The mutation in *Col3a1*^{Tsk2} substitutes a serine for the cysteine in one of the five Col 1-domain cysteines involved in disulfide bonds (Bruckner *et al.*, 1978).

Features shared by Tsk2/+ mice and people with fibrotic diseases (scleroderma, liver fibrosis, and kidney fibrosis) include the dysregulation of PIIINP (Sondergaard *et al.*, 1997; Majewski *et al.*, 1999; Abignano and Del Galdo, 2014; Del Galdo and Matucci-Cerinic, 2014; Quillinan *et al.*, 2014). The PIIINP fragment is a clinically validated biomarker of liver fibrosis (Leroy *et al.*, 2004; Rosenberg *et al.*, 2004) and scleroderma (Sondergaard *et al.*, 1997; Majewski *et al.*, 1999), and it has been used as a surrogate marker of fibrosis in clinical trials of potential SSc therapies (Majewski *et al.*, 1999; Denton *et al.*, 2009). Our finding of a point mutation in the protein that likely has a deleterious effect on PIIINP function is consistent with these clinical results and the fibrotic phenotype in the Tsk2/+ mouse.

Its high level in the sera of such patients may not merely be a benign biomarker. Support for this hypothesis derives from our *in vitro* complementation results showing that the presence of COL3A1-C33S is sufficient to upregulate the synthesis and secretion of COL1A1, consistent with the increased activity of the *Col1a1* promoter and excess production of COL1A1 in Tsk2/+ mice (Christner *et al.*, 1998; Barisic-Dujmovic *et al.*, 2008). It is likely that higher levels of or altered COL3A1 protein or PIIINP fragment also directly influence the composition and size of COL1A1/A2- and COL3A1-containing fibers, and that these features indirectly upregulate transforming growth factor- β 1 signaling, an important mediator of collagen production. A previous report from our laboratory has demonstrated increased dermal elastic fibers and transforming growth factor- β 1 accumulation in the skin of Tsk2/+ mice beginning at 2 weeks of age, lending further support to our hypothesis (Long *et al.*, 2014). In addition, our gene expression analyses show that similar global impact of the *Col3a1*^{Tsk2} gene occurs both *in vitro* and *in vivo*, and in both settings there are fundamental changes in the ECM and in fibroblasts due to the presence of this mutation. The hypothesis that *Col3a1*^{Tsk2} (or PIIINP^{Tsk2}) directly causes dermal fibrosis and scleroderma-like charac-

teristics is attractive: it would likely be dominant within the heterozygote, as collagen III is a homotrimeric triple helix (Ramachandran and Kartha, 1955), and the gene product of the mutant chromosome could be expected to contribute to alteration of a majority of collagen helices even in the presence of 50% normal collagen (Strachan and Read, 1999).

MATERIALS AND METHODS

All studies and procedures were approved by the Institutional Animal Care and Use Committee at Drexel University College of Medicine and conducted in accord with recommendations in the “Guide for the Care and Use of Laboratory Animals” (Institute of Laboratory Animal Resources, National Research Council, National Academy of Sciences). Detailed methods are provided in the Supplementary Materials online.

Animals

Tsk2/+ mice were serially backcrossed to the C57Bl/6J (B6) background. Recombinant B6.Tsk2/+ mice were also bred to B6.chr 1-A/J mice (Jackson Laboratory, Bar Harbor, ME) and the resulting B6.Tsk2/+ F1 mice were backcrossed to B6.chr 1-A/J mice. Wild-type littermates were used as controls.

DNA isolation from tail snips, microsatellite, and SNP typing

These were performed as in our previous publications (Bunce *et al.*, 1995; Butterfield *et al.*, 1998). Specific locations of SNP polymorphisms between B6 (which is very similar to 101/H) and A/J were determined using Mouse Genome Informatics (www.informatics.jax.org) and Mouse Phenome Database (<http://phenome.jax.org/>).

Complementation analysis

Tsk2/+ mice were crossed to *Col3a1* −/+ mice and their progeny mated to verify that the SNP in *Col3a1* is Tsk2. The resulting generations of the cross were genotyped by PCR for Tsk2/+ using microsatellites and primers specific to *Col3a1* or the inserted neomycin cassette (see Supplementary Material online).

In vitro assessment of fibrogenesis by COL3A1^{Tsk2}

We constructed a plasmid harboring the *Col3a1*^{Tsk2} allele by introducing the Tsk2 T-to-A mutation into a wild-type *Col3a1* clone (pCMV6-Kan/Neo; OriGene, Rockville, MD). A *Col3a1*-KO line was transfected with either plasmid as described (Artlett *et al.*, 1998). Supernatants were retained and cell lysates were harvested directly from the dish at 48 hours.

RNA isolation and real-time PCR

RNA was isolated from the skin or fibroblasts using a RNA isolation kit from Clontech (Mountain View, CA), and cDNAs synthesized from 2.0 μ g of total RNA using an High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA). Relative quantification of all products was measured using SYBR Green chemistry (Applied Biosystems).

RNA sequencing

Total RNA was prepared from three WT and four Tsk2/+ mice skin biopsies using the Qiagen RNeasy Fibrous Tissue Mini Kit (Qiagen Sciences, Germantown, MD). RNA-seq sequencing libraries were prepared for the seven samples using a NuGEN Ovation RNA-Seq

System (NuGEN Technologies, San Carlos, CA). Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform to obtain 16.7–50.9 million 50 bp paired-end reads per sample. The raw reads were aligned to the reference mouse genome (MM9 assembly) using Tophat software with default parameters (Trapnell *et al.*, 2009; Trapnell *et al.*, 2012). Supplementary Figure S1 online shows RNA-Seq read coverage for three interval genes. RNA-seq data from this study are available from NCBI Bioproject at accession number PRJNA262679.

454 Sequencing

Samples were captured and amplified as described in the Roche Nimblegen sequence capture manual (version 1.0; Madison, WI). Titanium general libraries were prepared from the captured DNAs from two 101/H mice and two Tsk2/+ mice using 5,000 ng of DNA. Enriched captured fragments were sequenced as described in GS FLX Titanium emPCR and Sequencing Protocols, October 2008. Sequence capture array probes were designed by Roche Nimblegen using the mouse genome sequence between 44,241,286 and 47,116,890 on chromosome 1 of mouse genome (MM9). Multiplexed 454 sequenced reads were assembled using Newbler v2.6 (454 Life Sciences, Branford, CT) with scaffolding against the same chromosome region that the probes were derived from.

DNA microarray hybridization and data analysis

This was performed as in our previous publications (Pendergrass *et al.*, 2012). RNA samples were amplified and labeled using the Agilent Low Input Linear Amplification kit (Agilent Technologies, Santa Clara, CA) and were hybridized against Universal Mouse Reference (Stratagene, La Jolla, CA) to Agilent Whole Mouse Genome arrays (G4122F; Agilent Technologies) in a common reference-based design. Microarrays were hybridized and washed in accordance with the manufacturer's protocols and scanned using a dual laser GenePix 4000B scanner (Axon Instruments, Foster City, CA). The pixel intensities of the acquired images were then quantified using GenePix Pro 5.1 software (Axon Instruments). Raw microarray data from this study are available from NCBI GEO at accession number GSE61728.

Western blot analyses

Culture supernatants were collected or the skin was homogenized in RIPA buffer (Sigma-Aldrich, St Louis, MO) using a glass homogenizer. Total protein was measured with a Bradford assay (Sigma-Aldrich), and western blots were performed as in our publications (Sassi-Gaha *et al.*, 2010). Antibodies used included goat anti-COL3A1 (#sc-8781), goat anti-COL1A1 (#sc-28657) from Santa Cruz Biotechnology, Santa Cruz, CA, rabbit anti- β -Actin (#4967, Cell Signaling Technologies, Boston, MA), donkey anti-goat (#705-035-003, Jackson ImmunoResearch Laboratories, West Grove, PA), or goat anti-rabbit (#111-035-003, Jackson ImmunoResearch), and signals were developed using SuperSignal West Dura ECL reagent (Thermo Scientific, Rockford, IL). Band intensities were measured using ImageQuant TL Software (GE Healthcare Life Sciences, Pittsburgh, PA).

Reticular fiber staining

Reticular fibers were stained using the Chandler's Precision Reticular Fiber Stain kit (American Master*Tech, Lodi, CA) according to the manufacturer's protocol.

Statistics

A two-tailed Student *t*-test or a one-way analysis of variance was used to determine statistical significance of collagen protein expression, as noted.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We thank Paul Christner for providing the breeding pairs of the original Tsk2/+ mice and Xianhua Piao at Harvard University for the *Col3a1*-KO mice. This work was supported by a Scleroderma Foundation Grant and awards from the National Institutes of Health (AR061384) and the Department of Defense (PR100338).

Author contributions

KBL and CMB bred and genotyped the B6.Tsk2 mice and all the derivative animals in this report; KBL, CMA, CMB, and SS-G conducted the histology on the skin and transfections on fibroblasts; EPB was responsible for the design and interpretation of the research including the genetic analyses; ZL and MLW conducted the expression analyses and interpreted the results; VM conducted GSEA analysis, SGC constructed the plasmids containing the mutant *Col3a1* cDNA; GDE, JE, RE, and AA performed the genomic DNA capture and sequencing and interpreted these results; KBL, EPB, CMA, and MLW wrote the paper.

SUPPLEMENTARY MATERIAL

Supplementary Material is linked to the online version of the paper at <http://www.nature.com/jid>

REFERENCES

- Abignano G, Del Galdo F (2014) Quantitating skin fibrosis: innovative strategies and their clinical implications. *Curr Rheumatol Rep* 16:404
- Artlett CM, Chen SJ, Varga J *et al.* (1998) Modulation of basal expression of the human alpha1(I) procollagen gene (COL1A1) by tandem NF-1/Sp1 promoter elements in normal human dermal fibroblasts. *Matrix Biol* 17:425–34
- Artlett CM (2010) Animal models of scleroderma: fresh insights. *Curr Opin Rheumatol* 22:677–82
- Asano Y, Stawski L, Hant F *et al.* (2010) Endothelial Flt1 deficiency impairs vascular homeostasis: a role in scleroderma vasculopathy. *Am J Pathol* 176:1983–98
- Barisic-Dujmovic T, Boban I, Clark SH (2008) Regulation of collagen gene expression in the Tsk2 mouse. *J Cell Physiol* 215:464–71
- Bruckner P, Bachinger HP, Timpl R *et al.* (1978) Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix leads to and comes from coil transition. *Eur J Biochem* 90:595–603
- Bunce M, O'Neill CM, Barnardo MC *et al.* (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46:355–67
- Butterfield RJ, Sudweeks JD, Blankenhorn EP *et al.* (1998) New genetic loci that control susceptibility and symptoms of experimental allergic encephalomyelitis in inbred mice. *J Immunol* 161:1860–7
- Christner PJ, Hitraya EG, Peters J *et al.* (1998) Transcriptional activation of the alpha1(I) procollagen gene and up-regulation of alpha1(I) and alpha1(III) procollagen messenger RNA in dermal fibroblasts from tight skin 2 mice. *Arthritis Rheum* 41:2132–42
- Christner PJ, Peters J, Hawkins D *et al.* (1995) The tight skin 2 mouse. An animal model of scleroderma displaying cutaneous fibrosis and mononuclear cell infiltration. *Arthritis Rheum* 38:1791–8
- Christner PJ, Siracusa LD, Hawkins DF *et al.* (1996) A high-resolution linkage map of the tight skin 2 (Tsk2) locus: a mouse model for scleroderma (SSc) and other cutaneous fibrotic diseases. *Mamm Genome* 7:610–2
- Cordes SP (2005) N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol Mol Biol Rev* 69:426–39

- Del Galdo F, Matucci-Cerinic M (2014) The search for the perfect animal model discloses the importance of biological targets for the treatment of systemic sclerosis. *Ann Rheum Dis* 73:635–6
- Denton CP, Engelhart M, Tvede N *et al.* (2009) An open-label pilot study of infliximab therapy in diffuse cutaneous systemic sclerosis. *Ann Rheum Dis* 68:1433–9
- Gentiletti J, McCloskey LJ, Artlett CM *et al.* (2005) Demonstration of autoimmunity in the tight skin-2 mouse: a model for scleroderma. *J Immunol* 175:2418–26
- Hocher B, Schwarz A, Fagan KA *et al.* (2000) Pulmonary fibrosis and chronic lung inflammation in ET-1 transgenic mice. *Am J Resp Cell Mol Biol* 23:19–26
- Leistritz DF, Pepin MG, Schwarze U *et al.* (2011) COL3A1 haploinsufficiency results in a variety of Ehlers-Danlos syndrome type IV with delayed onset of complications and longer life expectancy. *Genet Med* 13:717–22
- Leroy V, Monier F, Bottari S *et al.* (2004) Circulating matrix metalloproteinases 1, 2, 9 and their inhibitors TIMP-1 and TIMP-2 as serum markers of liver fibrosis in patients with chronic hepatitis C: comparison with PIIINP and hyaluronic acid. *Am J Gastroenterol* 99:271–9
- Liu X, Wu H, Byrne M *et al.* (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci USA* 94:1852–6
- Long KB, Artlett CM, Blankenhorn EP (2014) Tight skin 2 mice exhibit a novel time line of events leading to increased extracellular matrix deposition and dermal fibrosis. *Matrix Biol* 38:91–100
- Majewski S, Wojas-Pelc A, Malejczyk M *et al.* (1999) Serum levels of soluble TNF alpha receptor type I and the severity of systemic sclerosis. *Acta Derm Venereol* 79:207–10
- Maurer B, Busch N, Jungel A *et al.* (2009) Transcription factor fos-related antigen-2 induces progressive peripheral vasculopathy in mice closely resembling human systemic sclerosis. *Circulation* 120:2367–76
- Nishiyama Y, Nejima J, Watanabe A *et al.* (2001) Ehlers-Danlos syndrome type IV with a unique point mutation in COL3A1 and familial phenotype of myocardial infarction without organic coronary stenosis. *J Intern Med* 249:103–8
- Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome* 11:478–83
- Pendergrass SA, Lemaire R, Francis IP *et al.* (2012) Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J Invest Dermatol* 132:1363–73
- Peters J, Ball ST (1986) Tight Skin 2 (Tsk2). *Mouse News Letters* 74:91–2
- Quillinan NP, McIntosh D, Vernes J *et al.* (2014) Treatment of diffuse systemic sclerosis with hyperimmune caprine serum (AIMSPRO): a phase II double-blind placebo-controlled trial. *Ann Rheum Dis* 73:56–61
- Ramachandran GN, Kartha G (1955) Structure of collagen. *Nature* 176:593–5
- Richard V, Solans V, Favre J *et al.* (2008) Role of endogenous endothelin in endothelial dysfunction in murine model of systemic sclerosis: tight skin mice 1. *Fundam Clin Pharmacol* 22:649–55
- Rosenberg WM, Voelker M, Thiel R *et al.* (2004) Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 127:1704–13
- Sassi-Gaha S, Loughlin DT, Kappler F *et al.* (2010) Two dicarbonyl compounds, 3-deoxyglucosone and methylglyoxal, differentially modulate dermal fibroblasts. *Matrix Biol* 29:127–34
- Siracusa LD, McGrath R, Ma Q *et al.* (1996) A tandem duplication within the fibrillin 1 gene is associated with the mouse tight skin mutation. *Genome Res* 6:300–13
- Sondergaard K, Heickendorff L, Risteli L *et al.* (1997) Increased levels of type I and III collagen and hyaluronan in scleroderma skin. *Br J Dermatol* 136:47–53
- Strachan T, Read AP (1999) *Human Molecular Genetics*. 2nd edn. Wiley-Liss: New York
- Trapnell C, Pachter L, Salzberg SL *et al.* (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–11
- Trapnell C, Roberts A, Goff L *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–78
- Yamamoto T, Takagawa S, Katayama I *et al.* (1999) Animal model of sclerotic skin. I: local injections of bleomycin induce sclerotic skin mimicking scleroderma. *J Invest Dermatol* 112:456–62

Arthritis Research & Therapy

Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis --Manuscript Draft--

Manuscript Number:																
Full Title:	Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis															
Article Type:	Research article															
Section/Category:	Immunology and Pathology															
Funding Information:	<table border="1"> <tr> <td>National Institute of Arthritis and Musculoskeletal and Skin Diseases (1-P50-AR-060780-01)</td> <td>Dr. Michael L Whitfield</td> </tr> <tr> <td>U.S. Department of Defense (PR130908)</td> <td>Dr. Michael L Whitfield</td> </tr> <tr> <td>Dartmouth College (Synergy grant)</td> <td>Dr. Michael L Whitfield</td> </tr> <tr> <td>National Institute of Arthritis and Musculoskeletal and Skin Diseases (T32GM008704)</td> <td>Ms Jaclyn N Taroni</td> </tr> <tr> <td>National Institutes of Health (R01-CA155260)</td> <td>Dr. Scott A Gerber</td> </tr> <tr> <td>National Institutes of Health (S10-OD016212)</td> <td>Dr. Scott A Gerber</td> </tr> <tr> <td>Kellen Foundation (Clinician Scientist Development Award)</td> <td>Dr. Jessica K Gordon</td> </tr> </table>		National Institute of Arthritis and Musculoskeletal and Skin Diseases (1-P50-AR-060780-01)	Dr. Michael L Whitfield	U.S. Department of Defense (PR130908)	Dr. Michael L Whitfield	Dartmouth College (Synergy grant)	Dr. Michael L Whitfield	National Institute of Arthritis and Musculoskeletal and Skin Diseases (T32GM008704)	Ms Jaclyn N Taroni	National Institutes of Health (R01-CA155260)	Dr. Scott A Gerber	National Institutes of Health (S10-OD016212)	Dr. Scott A Gerber	Kellen Foundation (Clinician Scientist Development Award)	Dr. Jessica K Gordon
National Institute of Arthritis and Musculoskeletal and Skin Diseases (1-P50-AR-060780-01)	Dr. Michael L Whitfield															
U.S. Department of Defense (PR130908)	Dr. Michael L Whitfield															
Dartmouth College (Synergy grant)	Dr. Michael L Whitfield															
National Institute of Arthritis and Musculoskeletal and Skin Diseases (T32GM008704)	Ms Jaclyn N Taroni															
National Institutes of Health (R01-CA155260)	Dr. Scott A Gerber															
National Institutes of Health (S10-OD016212)	Dr. Scott A Gerber															
Kellen Foundation (Clinician Scientist Development Award)	Dr. Jessica K Gordon															
Abstract:	<p>Objective: Autoantibody profiles represent important patient stratification markers in systemic sclerosis (SSc). Here, we performed serum-immunoprecipitations with patient antibodies followed by mass spectrometry (LC-MS/MS) to obtain an unbiased view of all possible autoantibody targets and their associated molecular complexes recognized by SSc.</p> <p>Methods: HeLa whole cell lysates were immunoprecipitated (IP) using sera of patients with SSc clinically positive for autoantibodies against RNA polymerase III (RNAP3), topoisomerase 1 (TOP1), and centromere proteins (CENP). IP eluates were then analyzed by LC-MS/MS to identify novel proteins and complexes targeted in SSc. Target proteins were examined using a functional interaction network to identify major macromolecular complexes, with direct targets validated by IP-Western blots and immunofluorescence.</p> <p>Results: A wide range of peptides were detected across patients in each clinical autoantibody group. Each group contained peptides representing a broad spectrum of proteins in large macromolecular complexes, with significant overlap between groups. Network analyses revealed significant enrichment for proteins in RNA processing bodies (PB) and cytosolic stress granules (SG) across all SSc subtypes, which were confirmed by both Western blot and immunofluorescence.</p> <p>Conclusions: While strong reactivity was observed against major SSc autoantigens, such as RNAP3 and TOP1, there was overlap between groups with widespread reactivity seen against multiple proteins. Identification of PB and SG as major targets of the humoral immune response represents a novel SSc autoantigen and suggests a model in which a combination of chronic and acute cellular stresses result in aberrant cell death, leading to autoantibody generation directed against macromolecular nucleic acid-protein complexes.</p>															
Corresponding Author:	Michael E Johnson															

	Dartmouth College Geisel School of Medicine Hanover, NH UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Dartmouth College Geisel School of Medicine
Corresponding Author's Secondary Institution:	
First Author:	Michael E Johnson
First Author Secondary Information:	
Order of Authors:	Michael E Johnson
	Andrew V Grassetti
	Jaclyn N Taroni
	Shawn M Lyons
	Devin Schweppe
	Jessica K Gordon
	Robert F Spiera
	Robert Lafyatis
	Paul J Anderson
	Scott A Gerber
	Michael L Whitfield
Order of Authors Secondary Information:	
Opposed Reviewers:	

Click here to view linked References

Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis

Michael E. Johnson¹, Andrew V. Grassetti¹, Jaclyn N. Taroni¹, Shawn M. Lyons², Devin Schweppe¹, Jessica K. Gordon³, Robert F. Spiera³, Robert Lafyatis⁴, Paul J. Anderson², Scott A. Gerber¹, Michael L. Whitfield¹

Author Affiliations:

¹ Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

² Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA

³ Department of Rheumatology, Hospital for Special Surgery, New York, NY, USA

⁴ Boston University School of Medicine, Boston, MA, USA

Contact information:

MEJ: michael.e.johnson@dartmouth.edu; Dartmouth Medical School, Hinman Box 7400, Hanover, NH 03755.

AVG: andrew.v.grassetti.gr@dartmouth.edu; JNT: jaclyn.n.taroni.gr@dartmouth.edu; SML: smlyons@partners.org; dkschwep@uw.edu; JKG: gordonj@hss.edu; RFS: spierar@hss.edu; RL: rlafyatis@gmail.com; PJA: panderson@rics.bwh.harvard.edu; SAG: scott.a.gerber@dartmouth.edu; MLW: michael.l.whitfield@dartmouth.edu

ABSTRACT

Objective: Autoantibody profiles represent important patient stratification markers in systemic sclerosis (SSc). Here, we performed serum-immunoprecipitations with patient antibodies followed by mass spectrometry (LC-MS/MS) to obtain an unbiased view of all possible autoantibody targets and their associated molecular complexes recognized by SSc.

Methods: HeLa whole cell lysates were immunoprecipitated (IP) using sera of patients with SSc clinically positive for autoantibodies against RNA polymerase III (RNAP3), topoisomerase 1 (TOP1), and centromere proteins (CENP). IP eluates were then analyzed by LC-MS/MS to identify novel proteins and complexes targeted in SSc. Target proteins were examined using a functional interaction network to identify major macromolecular complexes, with direct targets validated by IP-Western blots and immunofluorescence.

Results: A wide range of peptides were detected across patients in each clinical autoantibody group. Each group contained peptides representing a broad spectrum of proteins in large macromolecular complexes, with significant overlap between groups. Network analyses revealed significant enrichment for proteins in RNA processing bodies (PB) and cytosolic stress granules (SG) across all SSc subtypes, which were confirmed by both Western blot and immunofluorescence.

Conclusions: While strong reactivity was observed against major SSc autoantigens, such as RNAP3 and TOP1, there was overlap between groups with widespread reactivity seen against multiple proteins. Identification of PB and SG as major targets of the humoral immune response represents a novel SSc autoantigen and suggests a model in which a combination of chronic and acute cellular stresses result in aberrant cell death, leading to autoantibody generation directed against macromolecular nucleic acid-protein complexes.

Keywords: Systemic sclerosis, scleroderma, autoantibody, RNA processing bodies, stress granules

Introduction

Systemic sclerosis (SSc) is a rare systemic autoimmune disease of unknown etiology characterized by skin fibrosis, internal organ involvement, vascular abnormalities, and autoantibody production. Patients are broadly classified as having either limited (lSSc) or diffuse (dSSc) disease based primarily upon the extent of skin involvement and autoantibody profiles. While a wide array of autoantibodies have been described for SSc, only a small number of these targets are used for clinical diagnosis and stratification. Autoantibodies targeting RNA polymerase III (RNAP3), topoisomerase 1 (TOP1; commonly referred to as Scl70), and centromere proteins (CENP) represent the three the most common, clinically measured autoantibodies observed in SSc [1, 2]. Other autoantibodies, including fibrillarin (U3RNP), Pm/Scl, Ku, U1RNP, U11/U12, and Th/To have also been described [1, 3] but are not routinely measured for clinical subtyping.

While the processes underlying autoantibody production in SSc remain poorly understood, the presence of certain autoantibodies is strongly predictive of clinical outcomes [1-3]. TOP1 and RNAP3 autoantibodies are almost exclusively seen in dSSc, while CENP, Th/To, and U1RNP antibodies are more commonly associated with lSSc [1, 3]. U3RNP autoantibodies are not associated with either clinical subset, and are often found in conjunction with other autoantibodies, including both TOP1 and CENP [3]. Certain antibodies, such as TOP1 and U11/12, have been shown to be predictive of poorer overall prognosis, including increased likelihood of pulmonary fibrosis [4] and cardiac involvement, while RNAP3 autoantibodies have recently been linked to co-occurrence of SSc with cancer [5].

Despite the importance of autoantibodies in SSc, the vast majority of target identification and phenotypic screening has been performed using methods targeting only a single autoantibody, with little ability to detect novel or low abundance autoantibodies. Furthermore, these methods fail to address the possibility of co-occurrence of multiple autoantibodies within a patient, which may have important clinical implications. Autoantigen microarrays have proven successful for screening large numbers of autoantibodies in parallel, however target identification is limited to those antigens produced and printed on the antigen microarrays [6]. To address these limitations, we performed immunoprecipitations (IP) of HeLa whole cell lysates using sera from RNAP3-, CENP-, and TOP1-positive patients, as well as healthy controls, followed by mass spectrometry (LC-MS/MS) to provide an unbiased assessment of all autoantibodies present in these SSc

1
2
3
4 patients. This method provides a better view of the full range of autoantibodies present in SSc,
5 including both novel and established targets, and provides insights into the general processes
6 underlying autoantibody production.
7
8
9

10 11 **Materials and Methods**

12 13 **Clinical Samples**

14
15 Patient serum was obtained from Boston University Medical School, Boston, MA, and the
16 Hospital for Special Surgery, New York, NY. All relevant study protocols were approved by
17 Dartmouth College's committee for the protection of human subjects, and the internal review
18 boards of both BUMC and HSS. Informed consent was obtained from all patients prior to
19 sample collection. All patients met the clinical classifications for either diffuse or limited SSc,
20 as set forth by the American College of Rheumatology. Diagnoses of major autoantibody
21 profiles were performed using standard clinical assays.
22
23
24
25
26
27
28
29

30 31 **Human Cell Lysates**

32 HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) (v/v)
33 and 100 IU/mL penicillin-streptomycin. Cells were grown to ~80% confluence, harvested in IP
34 lysis buffer (150 mM NaCl, 50 mM Tris pH 7.5, 1mM MgCl₂, 1mM EDTA, 0.5% Triton X-100,
35 2.5 mM β -mercaptoethanol, 1mM sodium molybdate, 1mM sodium fluoride, 1mM sodium
36 tartrate, 1 mM dithiothreitol (DTT), and protease inhibitors (Roche, Indianapolis, IN, USA)),
37 lysed by passage through a pre-chilled high-gauge syringe, and centrifuged for 15 min to pellet
38 debris. Lysates were then clarified by incubating for 4 h at 4°C on a rotating platform. Protein
39 concentrations were quantified using a standard BCA protein assay kit (Thermo Scientific,
40 Waltham, MA, USA).
41
42
43
44
45
46
47
48
49

50 51 **Serum Immunoprecipitation**

52 Patient serum was cross-linked to Protein G Dynabeads (Invitrogen, St. Louis, MO, USA)
53 prior to IP. First, 100 μ L serum (~1 mg IgG) was added to 50 μ L Protein G beads and incubated
54 for 5 h at 4°C. Samples were then washed in PBS, equilibrated in crosslinking buffer (50 mM
55 HEPES, pH 8.2), and cross-linked to Protein G beads by the addition of DMP solution (20 mM
56 dimethyl pimelimidate, 300 mM HEPES) for 10 min at room temperature (repeated three times).
57
58
59
60
61
62
63
64
65

1
2
3
4 The crosslinking reaction was then terminated by the addition of 50 mM ammonium bicarbonate,
5 and the resulting antibody bead mixture added to 500 μ L cell lysate (diluted to 4 mg/mL in IP
6 lysis buffer). Samples were incubated overnight at 4°C on a rotating platform, washed in cold IP
7 lysis buffer, and eluted in a buffer containing 2% SDS, 75 mM NaCl, 50 mM Tris pH 8.1, and
8 20% glycerol at 65°C for 5 min. Eluates were reduced by the addition of 0.1 M DTT (to a final
9 concentration 5 mM), and incubated at 80°C for 5 min. Samples were then resolved by SDS-
10 PAGE, split into high (> 60 kDa) and low (< 60 kDa) molecular weight fractions and analyzed
11 by mass spectrometry.
12
13
14
15
16
17
18
19
20

21 **Mass Spectrometry**

22 Proteins contained in Coomassie stained gel regions were digested overnight with trypsin
23 (1:200 w/v) at 37°C. Following digestion, peptides were extracted from the gels, dried, and
24 analyzed by nanoscale LC-MS/MS. LC-MS/MS analyses were performed on either LTQ
25 Orbitrap Classic or Orbitrap Fusion LC-MS/MS platforms. LTQ Orbitrap Classic analyses were
26 conducted as described previously [7].
27
28
29
30

31 For Orbitrap Fusion analyses, samples were loaded onto an EASY-nLC 1000 Liquid
32 Chromatograph (Thermo Scientific, Waltham, MA) and separated by reverse-phase high
33 pressure liquid chromatography (RP-HPLC) using a ~36 cm column with a 100 μ M inner
34 diameter packed with 3 μ m 120 Å C₁₈ particles (Dr. Maisch GmbH, Ammerbuch-Entringen,
35 Germany). The resultant peptide eluate was directed into an Orbitrap Fusion Tribrid Mass
36 Spectrometer operating in a data-dependent sequencing acquisition mode across a 30 min
37 reverse-phase gradient (6% acetonitrile, 0.1% formic acid to 30% acetonitrile, 0.1% formic acid)
38 at 350 nL/min flow rate. The Orbitrap Fusion was operated with an Orbitrap MS1 scan at 120K
39 resolution, followed by Orbitrap MS2 scans of higher energy collision induced dissociation
40 (HCD) fragment ions (30% HCD energy) at 15K resolution using a maximum cycle type of 2s,
41 precursor ion dynamic exclusion window of 15s, +2, +3, and +4 precursor ions selected for LC-
42 MS/MS, and maximum ion injection times of 100 ms (MS1) and 50 ms (MS2). The resulting
43 tandem mass spectra were data-searched using the COMET search engine [8] against a *Homo*
44 *sapiens* proteome database (Source: Uniprot, download date: 02-07-2013) with a precursor ion
45 tolerance of +/- 1Da [9] and a fragment ion tolerance of 0.02 Thomsons. Peptide spectra
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 matches (PSMs) were filtered to a $< 1\%$ false discovery rate using the target decoy strategy [10]
5 and reported.
6
7

8 9 10 **IP-Western Blots**

11 Anti-UPF1 antibody was kindly provided by Dr. Lynne Maquat (University of Rochester
12 Medical Center, Rochester, NY, USA). Antibodies to MOV10 and CAPRIN1 were purchased
13 from Proteintech (Chicago, IL, USA); antibodies to G3BP1 and USP10 were purchased from
14 Santa Cruz Biotechnology (Santa Cruz, CA, USA). Serum immunoprecipitation of HeLa lysates
15 was performed as described above; 50% of each eluate (15 μ L) was then run on a 10% bis-tris
16 precast gel (Life Technologies, Carlsbad, CA, USA). HeLa whole cell lysate (100 μ g) was used
17 as a positive control; no loading control was performed due to the absence of viable targets
18 present in all IP eluates. Western blots were then run following standard protocols, and
19 visualized using Western Lightning ECL Pro or Ultra substrate (Perkin Elmer Inc., Waltham,
20 MA, USA), as necessary.
21
22
23
24
25
26
27
28
29
30
31

32 **Data Analysis**

33 Non-redundant peptide hits, defined as mass spectra mapping exclusively to a given peptide
34 fragment, were used for all downstream analyses. Pair-wise comparisons between samples were
35 performed by Fisher's exact test using a Bonferroni correction for multiple hypothesis testing.
36 Venn diagrams were generated using VENNY [11]. Network analysis was performed using the
37 Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT;
38 <http://giant.princeton.edu/>) global network [12] and visualized using Cytoscape [13].
39 Communities in the network were detected using fast-greedy modularity as implemented in
40 igraph. Functional annotation of individual communities was performed using g:Profiler [14].
41 Semi-quantitative enrichment of SSc-associated autoantibodies was determined using a binary
42 assessment of autoantibody presence or absence in a sample. Preferential enrichment in SSc was
43 defined as all proteins detected in $> 50\%$ of all patient samples at a frequency > 1.5 -fold relative
44 to controls. Enrichment of biological processes and cellular components was determined using
45 g:Profiler using the g:SCS threshold correction for multiple hypothesis testing and a functional
46 category size of ≤ 500 genes. Hierarchical clustering was performed using Cluster 3.0 [15], and
47 visualized using Java TreeView [16].
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Immunofluorescence

The day prior to the experiment, 10^5 U2OS cells were seeded onto 11 mm glass coverslips and allowed to attach overnight at 37°C/5% CO₂ in DMEM containing 10% FBS (Gibco). Cells were treated with 100 μ M sodium (meta)arsenite (Sigma Aldrich) for 1 hr to induce the formation of stress granules and then with 4% paraformaldehyde solution at room temperature for 15 min followed by blocking and permeabilization with 5% normal horse serum, 0.1% digitonin in Tris-buffered saline. Staining was performed with anti-eIF3b (Santa Cruz), anti-SK1-Hedls (Santa Cruz), and patient sera for 1 hr at room temperature. Secondary antibodies (anti-goat-Cy3, anti-mouse-Cy2, and anti-human-Cy5) were purchased from Jackson Labs and incubated at room temperature for 1 hr. Conventional fluorescence microscopy was performed using a microscope (model Elipse E800, Nikon) with epifluorescence optics with a digital camera (model CCD-SPOT RT; Diagnostic Instruments). Images were compiled using Adobe Photoshop software (CS6).

Results

Identification of proteins cross-reacting to serum antibodies

Immunoprecipitations (IP) of HeLa whole cell lysates were performed using sera obtained from 13 SSc patients and 4 healthy controls. HeLa cells were chosen based upon their consistent, high level of expression of a broad range of proteins from the human genome [17].

SSc patients were divided into three groups, TOP1, RNAP3, and CENP, as measured in a reference laboratory; clinical data for each patient are shown in Table 1. Immunoprecipitated proteins were analyzed by LC-MS/MS, and the resulting spectra aligned to the reference human proteome (UCSC version hg19). Data are presented in two ways; first to identify the total number of peptides which could be aligned to each protein (total hits), and second to identify all non-redundant peptides which mapped exclusively to a given protein (non-redundant hits). A complete list of all data can be found in Supplemental Table S1.

Exclusivity and co-occurrence of SSc autoantibodies

We observed a high degree of reproducibility between patients within their respective autoantibody groups (TOP1, RNAP3, and CENP; Figure 1). The greatest degree of overlap

1
2
3
4 between peptides was observed among RNAP3 patients (Figure 1A and C), with 420 proteins
5 (54.1%) detected in all four patients (Figure 1C). The remaining groups exhibited significant
6 overlap in 3 of 4 (TOP1) and 4 of 5 (CENP) patients, respectively (Figure 1A), along with a
7 single outlier that showed either higher (SSc 208; TOP1) or lower (SSc 226; CENP) total peptide
8 hits relative to other samples in these groups. Within TOP1, 111 proteins (14.2%) were detected
9 in all four patients (Figure 1D), while CENP exhibited 48 proteins (10.5%) common to all
10 patients (Figure 1C). The least overlap was seen in healthy controls, with only 40 proteins
11 (7.6%; Figure 1B) common across individuals.
12
13
14
15
16
17
18

19 Across all samples, 283 proteins (25.0%) were detected in at least one patient in each of the
20 four autoantibody groups (Figure 1E, Supplemental Table S2). Some of these proteins likely
21 represent background signals (serum albumin (ALB), β -tubulin (TUBB), and ribosomal
22 proteins), while others are considered specific to SSc despite trace level detection in controls.
23 For example, multiple SSc autoantibody targets, including Ku (XRCC5 and XRCC6),
24 Ro52/TRIM21, and nucleophosmin/B23 (NPM1) were present in this set of proteins. In contrast,
25 87 proteins (7.7%) were detected in all three SSc groups, but were absent in controls (Figure 1E;
26 Supplemental Table S2). Functional analyses of these proteins revealed strong enrichment of
27 proteins involved in oxidative stress responses and nucleic acid processing (Supplemental Table
28 S3B).
29
30
31
32
33
34
35
36

37 Of the 1130 non-redundant proteins identified, 473 (41.8%) were unique to a given
38 autoantibody group (Figure 1F); however, the vast majority of these proteins were exclusive to a
39 single patient, with only 111 (23.5%) detected in two or more patients. These results suggest a
40 wide-range of autoantibody responses within each of the clinical autoantibody groups beyond
41 what has already been described.
42
43
44
45
46

47 Among the major autoantibody groups, immunoprecipitation of RNAP3 was exclusive to the
48 RNAP3 group, with no RNAP3 peptides detected in any of the other samples (Table 2). In
49 contrast, TOP1 peptides were consistently highest among TOP1 patients, but were also detected
50 at low levels in all four RNAP3⁺ patients, as well as two controls (Table 2). As these patients
51 were negative for TOP1 autoantibodies by clinical testing, these results indicate a higher degree
52 of sensitivity for our IP/MS protocol compared to standard ELISA-based methods used
53 clinically. In contrast, CENP was only detected at low levels in the CENP group, likely because
54 it remained bound to the tightly packed centromere complex of chromatin.
55
56
57
58
59
60
61
62
63
64
65

Other known SSc autoantigens were also detected. RuvBL [18] was strongly detected in all SSc samples, while virtually absent in controls. Ku and Su, along with a wide array of anti-tRNA synthetases [19] were routinely detected in both the RNAP3 and TOP1 subsets, but were only weakly present in the CENP and control groups (Table 2).

Several autoantigens previously implicated in SSc were found at low, background levels in both SSc and control samples. Ro52/TRIM21 [20] and nucleophosmin/B23 [21] were widely detected across all four groups, suggesting a high degree of background reactivity to these proteins in SSc and controls. We did not find evidence for enrichment in SSc of Pm/Scl autoantibodies, which target exosome components EXOSC1-10 [22]. Peptides for these proteins were absent in the CENP group, but were detected at low levels in other subsets, including controls. Autoantigens not detected here include many of the URNPs, PDGFR, matrix metalloproteinases, tissue plasminogen activator, and vascular receptor antibodies (Table 2).

Functional clustering of identified proteins

To identify functional interactions among autoantigens, all 763 non-redundant protein hits were submitted as a query to the GIANT global average network. This approach included both SSc-specific targets as well as those detected at background levels in controls to better understand the full range of autoreactive proteins and complexes. Nine distinct communities were identified within the resulting network, in which each gene is represented by a node, and two genes share an edge if they are predicted to functionally interact (Supplemental Figure S1). Analysis of each of these communities by g:Profiler revealed functional enrichment for a wide range of biological processes associated with important disease processes and components (Supplemental Figure S1). Community 1 is dominated by ribosomal proteins, eukaryotic initiation factor 3 (eIF3) subunits, and includes the SSc autoantibody target nucleophosmin/B23. Communities 2 and 8 show strong enrichment for GO terms *mRNA processing*, *ribonucleoprotein complex*, and *cytosolic stress granule*. Community 2 is dominated primarily by DEAD box helicases proteins, while community 8 contains a diverse array of proteins including multiple SSc autoantibodies, including TOP1, SSB, Pm/Scl proteins, URNPs, and HNRNPs, as well as numerous serine/arginine-rich splicing factors. Community 3 consists primarily of aminoacyl tRNA synthetases, a cluster often targeted in autoimmune diseases [19, 23]. Communities 4, 5, and 9 are strongly associated with a variety of GO processes known to

1
2
3
4 play a major role in SSc, including *wound healing*, *IFN signaling*, and *response to oxidative*
5 *stress*. Major proteins include CD44, HLAs, myosins, and filamin proteins in community 4 and
6 tricarboxylic acid cycle proteins in community 5. Community 9 contains multiple protein
7 disulfide isomerases and peroxiredoxins, protein folding enzymes such as calnexin (CANX) and
8 calreticulin (CALR), and the major collagen processing enzyme prolyl 4-hydroxylase beta
9 (P4HB). Community 6 contains multiple annexin and 14-3-3 proteins; enriched GO processes
10 include *ribonucleoprotein complex assembly*, *mitochondrial transport*, *RNA processing*, and
11 *anchoring junction*. Community 7 associated with GO terms include *cell cycle*, *RNA polymerase*
12 *III complex*, *DNA-PK-Ku complex*, and *antigen processing and presentation*. Community 7
13 includes several SSc autoantibodies targets including Ku proteins XRCC5 and 6, RUVBL1 and
14 2, RNA polymerase I and II subunits, multiple proteasomal subunits, and T-complex proteins.
15
16
17
18
19
20
21
22
23
24
25

26 **Preferential detection of autoantibodies in SSc**

27
28 Subsequent comparisons between groups were performed in a semi-quantitative manner
29 based on the presence or absence of a given protein in an immunoprecipitant, with quantitative
30 analyses limited to comparisons within an individual sample. To identify biological processes
31 and cellular components differentially targeted in SSc, with minimal to no background detection
32 in controls, we examined all proteins detected in > 50% of SSc samples at a frequency > 1.5-fold
33 relative to controls, resulting in a list of 137 differentially detected proteins (Figure 2;
34 Supplemental Table S2). Enriched biological processes included *ncRNA metabolic process*,
35 *response to oxygen radical*, and *triglyceride-rich lipoprotein particle remodeling*. Preferentially
36 targeted cellular components include *cytosolic stress granule*, *lipid-protein complex*, *pigment*
37 *granule*, and *anchoring junction*; molecular functions include *antioxidant activity* and *mRNA*
38 *binding* (Supplemental Table S3C).
39
40
41
42
43
44
45
46
47
48
49

50 **RNA processing centers are major targets of SSc autoantibodies**

51
52 The strong enrichment for GO terms associated with mRNA processing and stress response,
53 as well as the identification of *cytosolic stress granule* as an enriched cellular component, led us
54 to further investigate the role of stress granules (SG) and RNA processing bodies (PB) in the
55 autoantibody response of SSc. SGs and PBs represent distinct, non-membranous cytoplasmic
56 entities which arise in response to different cellular stresses, including oxidative stress, hypoxia,
57
58
59
60
61
62
63
64
65

1
2
3
4 viral infection, unfolded proteins, and amino acid deprivation [24]. These structures exist in
5 constant flux, driven by the availability of constituent mRNPs, regulating the fate of untranslated
6 mRNAs in response to translational arrest [25]. While SGs are generally absent under normal
7 conditions, PBs are constitutively present at low levels due to their role as microRNA processing
8 centers. Both structures have been shown to arise in response to cellular stresses, including
9 oxidative stress, ischemia, and cancer [26], all of which are known to be important in SSc
10 pathogenesis [5, 27].

11
12 In addition to the 137 differentially detected proteins described above, a wide range of
13 PB/SG constituents were readily detected across most SSc samples (Supplemental Table S4).
14 Substantial reactivity was seen against PB components such as UPF1 and MOV10, as well as SG
15 proteins FXR1 and FXR2, G3BP1 and G3BP2, and USP10. Only background levels of
16 reactivity was seen in healthy controls.

17 18 **Validation of PB/SG antibodies in SSc**

19
20 In order to validate the differential abundance of PB/SG proteins identified by LC-MS/MS,
21 HeLa whole cell lysates were immunoprecipitated using antibodies from each patient as
22 described in the LC-MS/MS analyses. Western blots were performed by resolving equal
23 volumes of IP eluates by SDS-PAGE and transferring to nitrocellulose. Blots were then probed
24 with antibodies targeting PB/SG proteins UPF1, MOV10, CAPRIN1, G3BP1, and USP10.
25 Strong reactivity was seen against all five proteins in SSc with only background reactivity in
26 controls (Figure 3A), indicating widespread immune responses against these protein complexes.

27
28 Further validation was performed using immunofluorescence (IF) staining of U2OS cells
29 maintained under conditions of oxidative stress to induce PB/SG formation. Cells were probed
30 with patient sera in combination with PB and SG markers SK1-Hedls and eIF3b, respectively.
31 Co-localization between patient sera and PB/SG markers was observed in 6 of 9 SSc patients,
32 with at least one positive sample in each of the three autoantibody groups; no staining was seen
33 for any of the three healthy controls (Figure 3B). These results are consistent with that seen by
34 LC-MS/MS, particularly among RNAP3 patients, who exhibited the strongest and most
35 consistent autoantibody responses across both methods. Taken together, these data strongly
36 implicate PB/SG as novel targets of SSc autoantibody responses.

Discussion

Autoantibodies have long been used in the diagnosis of SSc, with different autoantibodies predictive of clinical outcomes, including interstitial lung disease, pulmonary arterial hypertension, and skin involvement. While a wide array of SSc-associated autoantibodies have been described, diagnoses are often performed based upon the presence or absence of reactivity against three proteins: RNAP3, TOP1, and CENP. The data presented here suggest a much broader autoantibody response, which is reflective of underlying disease pathologies. Strong subset-specific reactivity was evident against both RNAP3 and TOP1, with no RNAP3 peptides detected in any of the other groups; however, all four RNAP3 patients exhibited modest reactivity against TOP1, indicating a degree of background reactivity against this protein. When peptides recovered are extended beyond the three major targets, we find substantial overlap across the three major SSc groups. We find peptides from the autoantigens of RuvBL1/2, which appear to act as general markers of SSc, with consistent detection across all SSc groups, with almost no background reactivity seen in controls. In contrast, some common SSc autoantigens such as B23 and Ro52/TRIM21 were recovered in virtually all samples, as well as controls, indicating an important degree of baseline reactivity against some of the more common autoantibody targets.

In this proof-of-concept study, we do not attempt to address the clinical implications of the autoantibody responses described here due to the limited number of patients analyzed. Our depth in this study comes from the number of potential antigens analyzed, which cover the full proteome. Future studies examining a much larger cohort of SSc patients, along with representatives of other autoimmune diseases, will be necessary to determine the clinical value of these potential autoantibodies.

This is not the first study to suggest the presence of multiple autoantibodies in SSc. Immunoassays performed by Op De Beeck, *et al.* revealed the presence multiple autoantibodies in a small subset of SSc patients [28]. A similar analysis by Graf *et al.* using the EUROLINE immunoassay revealed the presence of multiple autoantibodies in 11% of patients [1].

Autoantibodies against extracellular immune signaling receptors and extracellular matrix proteins were conspicuously absent in these data; this includes the absence of numerous autoantibodies previously implicated in SSc pathogenesis, such as anti-fibrillin 1, anti-MMP, and anti-PDGFR [29]. Additional analyses in other cell types, such as fibroblasts or endothelial

1
2
3
4 cells, as well as cells maintained under physiologically relevant growth conditions, such as
5 immune activation or oxidative stress, may be useful for identifying other proteins and
6 complexes which may play a role in disease pathogenesis.
7

8
9 In addition to identifying novel autoantibody targets, the unbiased nature of mass
10 spectrometry provides additional insights into the processes potentially underlying
11 autoimmunity. The preferential detection of proteins associated with RNA processing and
12 oxidative stress as a general feature of SSc autoantibodies may be indicative of their origins.
13 Combined with the consistent targeting of PB/SG described here spanning all SSc patients, these
14 data suggest a basic model in which disease-specific pathologies give rise to specific
15 autoantibodies. Strong induction of SGs is observed in response to cellular stresses, including
16 oxidative stress and ischemia, two well-established phenomena in SSc [27]. SG/PB are also
17 readily induced in response to the tumor microenvironment, consistent with recent evidence
18 linking RNAP3-positive SSc and cancer [5, 30]. Combined with evidence linking TGF- β
19 signaling with an increase in PB formation [31], many of the major processes underlying SSc
20 pathogenesis appear broadly consistent with an immune response against cells undergoing a
21 stress response. PBs are also known to associate with other cytoplasmic structures, such as U
22 bodies [32], which house an number of well-established SSc autoantibody targets, including U1,
23 U5, and U11/U12. Taken together, these data suggest a model in which autoantibodies arise as a
24 secondary phenotype in response to SSc-related processes already underway. Comparison to
25 other rheumatic diseases will allow us to understand if reactivity to SG/PBs is a common feature
26 of autoimmune diseases.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

42 This work has several limitations. First, we cannot eliminate the possibility that some
43 proteins found in our mass spec data result from co-IP of multi-protein complexes by a single
44 autoantibody; however, we were able to confirm the presence of multiple PB/SG autoantibodies
45 by other means (Figure 3). We also cannot rule out the possibility that some targets were missed
46 due to their being sequestered into tightly packed molecular complexes associated with
47 chromatin. For example, the presence of CENP autoantibodies within these samples had been
48 established using clinical methods, indicating its absence in our mass spec data is likely a result
49 of its sequestration into large macromolecular complexes with limited solubility. The small
50 number of patient samples used in this study prevents any clinical interpretation, and the
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

variability in the number of peptides recovered between experiments limits direct quantitative comparisons between autoantibody groups.

Conclusions

The data presented here provide evidence of diverse immune reactivities in SSc targeting a wide array of protein complexes. Among these complexes, autoantibodies targeting PB/SG were consistently identified across both clinical SSc subsets and major autoantibody groups, suggesting a potential novel autoantibody target. Taken together, these data suggest immune responses to proteins involved in cellular stress may be a common mechanism for autoantibody generation.

Abbreviations:

SSc, systemic sclerosis; dSSc, diffuse systemic sclerosis; lSSc, limited systemic sclerosis; LC-MS/MS, liquid chromatography tandem-mass spectrometry, IP, immunoprecipitation; CENP, centromere protein; RNAP3; RNA polymerase III; TOP1, topoisomerase I; PB, RNA processing bodies; SG, stress granules

Acknowledgements

This work was supported by grants from the NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases [NIAMS] Center of Research Translation (1-P50-AR-060780-01 to MLW and RL), the Department of Defense (PR130908 to MLW), and a SYNERGY grant from the Geisel School of Medicine at Dartmouth (to MLW). JNT is supported in part by a grant from the National Institute of General Medical Sciences (NIGMS; T32GM008704). SAG is supported by grants R01-CA155260 and S10-OD016212 from the NIH. JKG is supported by a Kellen Foundation Clinician Scientist Development Award from the Hospital for Special Surgery. Written informed consent was obtained from all participants for publication of their individual details in this manuscript. All consent forms are held by the respective authors' institutions and are available for review by the Editor-in-Chief.

Conflicts of Interest

1
2
3
4 Dr. Whitfield has received royalties for patents regarding gene expression biomarkers in
5 Scleroderma and is a scientific founder of Celdara Medical LLC. Dr. Lafyatis has received both
6 grants and consulting fees from Genzyme/Sanofi, Shire, Regeneron, Biogen, BMS, Inception,
7 Precision Dermatology, PRISM, UCB, Pfizer and Roche/Genentech; he received consulting fees
8 from Lycera, Novartis, Celgene, Amira, Celdara, Celltex, Dart Therapeutics, Idera, Intermune,
9 Medimmune, Promedior, Zwitter, Actelion, EMD Serono, Akros, Extera, Reneo, Scholar Rock,
10 and HGS. No authors have any non-financial conflicts of interest to report.
11
12
13
14
15
16
17
18

19 **Authors' Contributions**

20 MEJ conceived of the study, performed experiments, analyzed data, and wrote the
21 manuscript. AVG performed mass spectrometry and helped to revise the manuscript. JNT
22 performed data analysis and helped to revise the manuscript. SML performed
23 immunofluorescence experiments, and helped to revise the manuscript. DS performed mass
24 spectrometry. JKG, RFS, and RL provided clinical samples. PJA designed experiments and
25 provided technical assistance. SAG and MLW conceived of the study, participated in its design,
26 and helped to revise the manuscript.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

1. Graf SW, Hakendorf P, Lester S, Patterson K, Walker JG, Smith MD, Ahern MJ, Roberts-Thomson PJ: **South Australian Scleroderma Register: autoantibodies as predictive biomarkers of phenotype and outcome.** *Int J Rheum Dis* 2012, **15**(1):102-109.
2. Steen VD: **Autoantibodies in systemic sclerosis.** In: *Sem Arth Rheum: 2005*: Elsevier; 2005: 35-42.
3. Mehra S, Walker J, Patterson K, Fritzler MJ: **Autoantibodies in systemic sclerosis.** *Autoimmunity Rev* 2013, **12**(3):340-354.
4. Fertig N, Domsic RT, Rodriguez-Reyna T, Kuwana M, Lucas M, Medsger TA, Feghali-Bostwick CA: **Anti-U11/U12 RNP antibodies in systemic sclerosis: A new serologic marker associated with pulmonary fibrosis.** *Arthritis Care Res* 2009, **61**(7):958-965.
5. Joseph CG, Darrah E, Shah AA, Skora AD, Casciola-Rosen LA, Wigley FM, Boin F, Fava A, Thoburn C, Kinde I: **Association of the autoimmune disease scleroderma with an immunologic response to cancer.** *Science* 2014, **343**(6167):152-157.
6. Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ, Fournel S, Fong D, Genovese MC, Neuman de Vegvar HE *et al*: **Autoantigen microarrays for multiplex characterization of autoantibody responses.** *Nature Med* 2002, **8**(3):295-301.
7. Yore MM, Kettenbach AN, Sporn MB, Gerber SA, Liby KT: **Proteomic analysis shows synthetic oleanane triterpenoid binds to mTOR.** *PLoS ONE* 2011, **6**(7):e22862.
8. Eng JK, Jahan TA, Hoopmann MR: **Comet: An open-source MS/MS sequence database search tool.** *Proteomics* 2013, **13**(1):22-24.
9. Hsieh EJ, Hoopmann MR, MacLean B, MacCoss MJ: **Comparison of database search strategies for high precursor mass accuracy MS/MS data.** *J Proteome Res* 2009, **9**(2):1138-1143.
10. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.** *Nat Methods* 2007, **4**(3):207-214.
11. Oliveros JC: **VENNY. An interactive tool for comparing lists with Venn Diagrams.** In.; 2007.
12. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC: **Understanding multicellular function and disease with human tissue-specific networks.** *Nat genet* 2015.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498-2504.
14. Reimand J, Arak T, Vilo J: **g: Profiler—a web server for functional interpretation of gene lists (2011 update).** *Nucleic Acids Res* 2011, **39**(suppl 2):W307-W315.
15. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *PNAS* 1998, **95**(25):14863-14868.
16. Saldanha AJ: **Java Treeview—extensible visualization of microarray data.** *Bioinformatics* 2004, **20**(17):3246-3248.
17. Novoradovskaya N, Perou C, Whitfield M, Basehore S, Pesich R, Aprelikova O, Fero M, Brown P, Botstein D, Braman J: **Universal human, mouse and rat reference RNA as standards for microarray experiments.** In: *Mol Bio Cell: 2002*: AMER SOC CELL BIOLOGY 8120 WOODMONT AVE, STE 750, BETHESDA, MD 20814-2755 USA; 2002: 241A-241A.
18. Kaji K, Fertig N, Medsger TA, Satoh T, Hoshino K, Hamaguchi Y, Hasegawa M, Lucas M, Schnure A, Ogawa F *et al*: **Autoantibodies to RuvBL1 and RuvBL2: A novel systemic sclerosis-related antibody associated with diffuse cutaneous and skeletal muscle involvement.** *Arthritis Care Res* 2013:n/a-n/a.

19. Hamaguchi Y, Fujimoto M, Matsushita T, Kaji K, Komura K, Hasegawa M, Kodera M, Muroi E, Fujikawa K, Seishima M *et al*: **Common and Distinct Clinical Features in Adult Patients with Anti-Aminoacyl-tRNA Synthetase Antibodies: Heterogeneity within the Syndrome.** *PLoS ONE* 2013, **8**(4):e60442.
20. Fujimoto M, Shimozuma M, Yazawa N, Kubo M, Ihn H, Sato S, Tamaki T, Kikuchi K, Tamaki K: **Prevalence and clinical relevance of 52-kDa and 60-kDa Ro/SS-A autoantibodies in Japanese patients with systemic sclerosis.** *Ann Rheum Dis* 1997, **56**(11):667-670.
21. Ulanet DB, Wigley FM, Gelber AC, Rosen A: **Autoantibodies against B23, a nucleolar phosphoprotein, occur in scleroderma and are associated with pulmonary hypertension.** *Arth Care Res* 2003, **49**(1):85-92.
22. Brouwer R, Vree Egberts WTM, Hengstman GJD, Raijmakers R, van Engelen BGM, Peter Seelig H, Renz M, Mierau R, Genth E, Pruijn GJM *et al*: **Autoantibodies directed to novel components of the PM/ScI complex, the human exosome.** *Arth Res* 2002, **4**(2):134-138.
23. Lega J-C, Fabien N, Reynaud Q, Durieu I, Durupt S, Dutertre M, Cordier J-F, Cottin V: **The clinical phenotype associated with myositis-specific and associated autoantibodies: A meta-analysis revisiting the so-called antisynthetase syndrome.** *Autoimmunity Rev* 2014, **13**(9):883-891.
24. Kedersha N, Ivanov P, Anderson P: **Stress granules and cell signaling: more than just a passing phase?** *Trends Biochem Sci* 2013, **38**(10):494-506.
25. Kedersha N, Anderson P: **Chapter 4 Regulation of Translation by Stress Granules and Processing Bodies.** In: *Prog Mol Biol Transl Sci*. Edited by John WBH, vol. Volume 90: Academic Press; 2009: 155-185.
26. Anderson P, Kedersha N: **Stress granules: the Tao of RNA triage.** *Trends in biochemical sciences* 2008, **33**(3):141-150.
27. Katsumoto TR, Whitfield ML, Connolly MK: **The pathogenesis of systemic sclerosis.** *Annu Rev Pathol-Mech* 2011, **6**:509-537.
28. Op De Beéck K, Vermeersch P, Verschueren P, Westhovens R, Mariën G, Blockmans D, Bossuyt X: **Antinuclear antibody detection by automated multiplex immunoassay in untreated patients at the time of diagnosis.** *Autoimmunity Rev* 2012, **12**(2):137-143.
29. Chung L, Utz P: **Antibodies in scleroderma: Direct pathogenicity and phenotypic associations.** *Curr Rheumatol Rep* 2004, **6**(2):156-163.
30. Anderson P, Kedersha N, Ivanov P: **Stress granules, P-bodies and cancer.** *Biochim Biophys Acta* 2014(0).
31. Blanco FF, Sanduja S, Deane NG, Blackshear PJ, Dixon DA: **Transforming growth factor β regulates P-body formation through induction of the mRNA decay factor tristetraprolin.** *Mol Cell Bio* 2014, **34**(2):180-195.
32. Liu J-L, Gall JG: **U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies.** *PNAS* 2007, **104**(28):11655-11659.
33. Dib H, Tamby MC, Bussone G, Regent A, Berezné A, Lafine C, Broussard C, Simonneau G, Guillevin L, Witko-Sarsat V *et al*: **Targets of anti-endothelial cell antibodies in pulmonary hypertension and scleroderma.** *Eur Respir J* 2012, **39**(6):1405-1414.
34. Naniwa T, Sugiura Y, Banno S, Yoshinouchi T, Matsumoto Y, Ueda R: **Ribosomal P protein P0 as a candidate for the target antigen of anti-endothelial cell antibodies in mixed connective tissue.** *Clin Exp Rheumatol* 2007, **25**:593-598.
35. Terrier B, Tamby MC, Camoin L, Guilpain P, Broussard C, Bussone G, Yaïci A, Hotellier F, Simonneau G, Guillevin L *et al*: **Identification of target antigens of antifibroblast antibodies in pulmonary arterial hypertension.** *Am J Resp Crit Care Med* 2008, **177**(10):1128-1134.

36. Terrier B, Tamby MC, Camoin L, Guilpain P, Bérézné A, Tamas N, Broussard C, Hotellier F, Humbert M, Simonneau G: **Anti-fibroblast antibodies from systemic sclerosis patients bind to α -enolase and are associated with interstitial lung disease.** *Ann Rheum Dis* 2009.
37. Satoh M, Chan JY, Ceribelli A, del-Mercado MV, Chan EK: **Autoantibodies to Argonaute 2 (Su antigen).** In: *Ten Years of Progress in GW/P Body Research*. Springer; 2013: 45-59.
38. Kubo M, Ihn H, Kuwana M, Asano Y, Tamaki T, Yamane K, Tamaki K: **Anti-U5 snRNP antibody as a possible serological marker for scleroderma–polymyositis overlap.** *Rheumatol* 2002, **41**(5):531-534.
39. Svegliati Baroni S, Santillo M, Bevilacqua F, Luchetti M, Spadoni T, Mancini M, Fraticelli P, Sambo P, Funaro A, Kazlauskas A *et al*: **Stimulatory autoantibodies to the PDGF receptor in systemic sclerosis.** *N Engl J Med* 2006, **354**(25):2667-2676.
40. Hashish L, Trieu E, Sadanandan P, Targoff I: **Identification of autoantibodies to tyrosyl-tRNA synthetase in dermatomyositis with features consistent with anti-synthetase syndrome.** In: *Arthritis Rheum: 2005*; 2005: S312-S312.
41. Betteridge Z, Gunawardena H, North J, Slinn J, McHugh N: **Anti-synthetase syndrome: a new autoantibody to phenylalanyl transfer RNA synthetase (anti-Zo) associated with polymyositis and interstitial pneumonia.** *Rheumatol* 2007, **46**(6):1005-1008.

Table Legends

Table 1. Clinical information of patients involved in this study. ANA = anti-nuclear antibody; HSS = Hospital for Special Surgery, New York, NY; BUMC = Boston University Medical Center, Boston, MA; ILD = interstitial lung disease; PAH = pulmonary arterial hypertension; MRSS = modified Rodnan skin score. Empty wells indicate information not available at the time of sample collection.

Table 2. SSc-associated autoantibodies observed in this study. Data are presented as the average of all peptide hits across each autoantibody group, followed by the frequency of peptide detection within the group. For autoantibodies known to target more than one protein or subunit, data for a single representative protein is shown, with the specific protein highlighted in bold. Associated proteins indicate specific protein targets identified in this study; among autoantibodies not identified here, the most common targets are listed. SSc, systemic sclerosis; lSSc, limited cutaneous SSc; dSSc, diffuse cutaneous SSc; PAH, pulmonary arterial hypertension; ILD, interstitial lung disease; CREST, CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia); PM/Scl, polymyositis/scleroderma; PM/DM, polymyositis/dermatomyositis. Symbols: -, +, ++, and +++ indicate an average of 0, 1 - 4, 5 - 9, and ≥ 10 peptide hits per group, respectively.

Figure Legends

Figure 1. Overview of mass spectrometry results. A) Correlation matrix of non-redundant protein hits for all patients and controls. Comparisons were performed using a Fisher's exact test with Bonferroni correction. Black boxes indicate intra-group comparisons for each of the four clinically-defined groups. Green = controls; Red = RNAP3; Blue = CENP; Yellow = TOP1. B-F) Venn diagrams depicting overlap in non-redundant peptide hits within and between groups. B) healthy controls, C) RNAP3, D) CENP, E) TOP1, and F) overlap between groups.

Figure 2. Proteins differentially detected in SSc. Semi-quantitative enrichment of SSc-associated autoantibodies was determined using a binary assessment of autoantibody presence or absence in a sample. Preferential enrichment in SSc was defined as all proteins detected in > 50% of all patient samples at a frequency > 1.5-fold relative to controls. A) Heat map of proteins differentially detected in SSc. B) Network analysis of differentially detected proteins. Community detection was performed using the GIANT global network; functional annotation was performed using gProfiler.

Figure 3. Validation of PB/SG as a target of the SSc autoimmune response. A) HeLa cell lysates were immunoprecipitated using patient sera, resolved by SDS-PAGE, and probed with antibodies targeting known PB and SG proteins; HeLa whole cell lysate was used as a control. B) Immunofluorescence was performed in U2OS cells treated with sodium (meta)arsenite to induce the formation of stress granules. Cells were then fixed with 4% paraformaldehyde and permeabilized with 5% normal horse serum and 0.1% digitonin in Tris-buffered saline. Staining was performed with anti-eIF3b (SG marker), anti-SK1-Hedls (PB marker), and patient sera. Representative images depicting co-localization between patient sera and SG/PB markers are shown, with sites of co-localization circled in red.

Supplemental Figure S1. Network analysis of SSc autoantigens. All 763 non-redundant peptide hits identified in 2 or more patients were analysis using the Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT) global network to identify functionally-associated protein

1
2
3
4 networks. Analysis of community function was performed using gProfiler. SSc-associated
5 autoantibodies are highlighted in yellow.
6
7
8

9
10 **Supplemental Table S1.** Complete list of peptides identified in this analysis. TP, number of
11 total peptides mapping to a protein; UP, number of unique peptides mapping to a protein; UM,
12 number of non-redundant peptides mapping exclusively to a protein; MW, molecular weight;
13 Length, protein length in amino acids.
14
15
16
17

18
19 **Supplemental Table S2.** SSc-specific enrichment of processes and components. Proteins
20 differentially detected in SSc were analyzed using gProfiler. Statistically significant processes
21 and components are shown. A) Peptides detected at any level across all four groups. B) Peptides
22 identified in all SSc groups, but absent in controls. C) Analysis of 137 proteins differentially
23 detected in SSc. BP, biological process; CC, cellular component; MF, molecular function; ke,
24 KEGG pathway; re, REACTOME pathway.
25
26
27
28
29
30

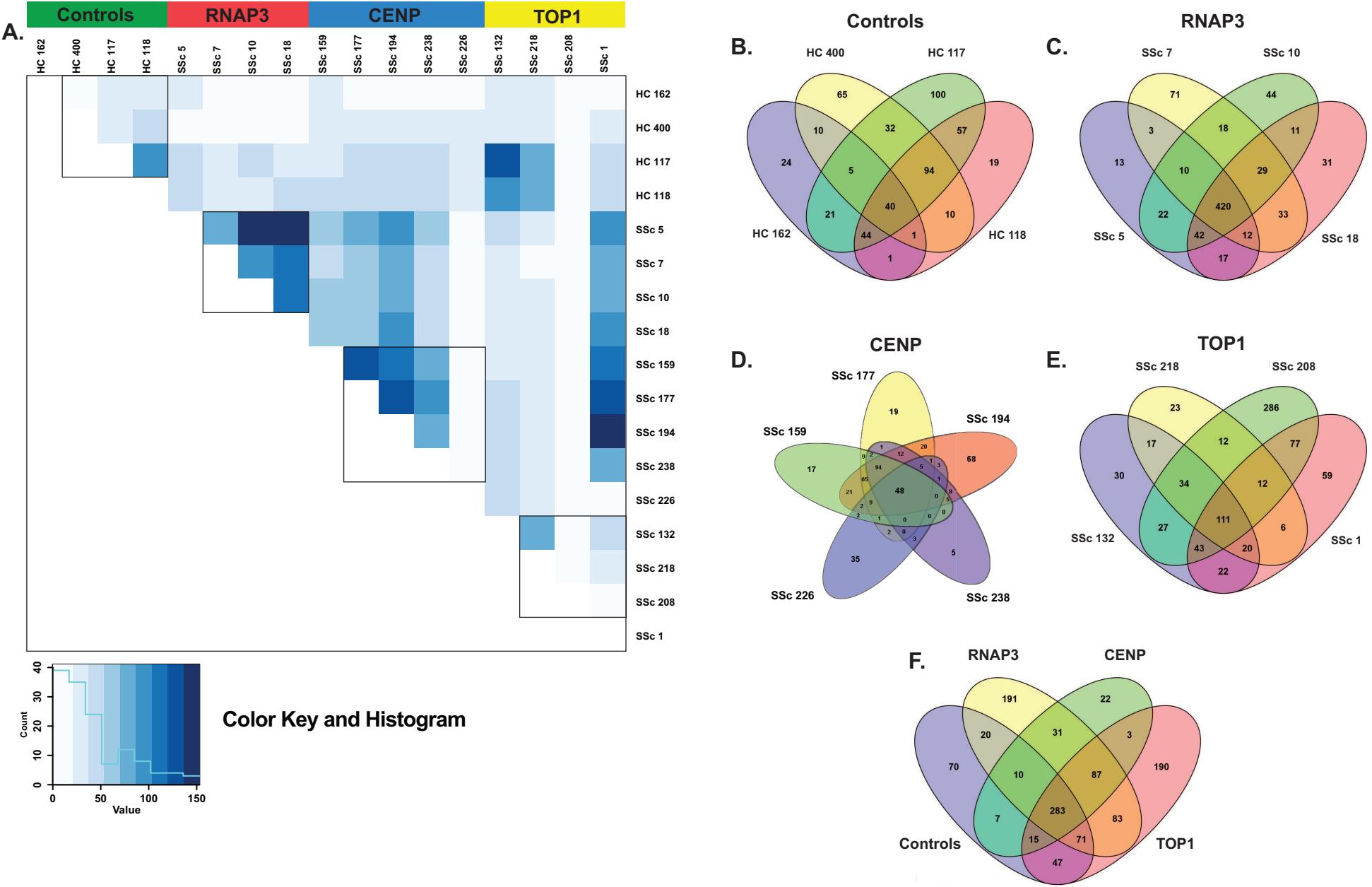
31
32 **Supplemental Table S3.** Processing body and stress granule proteins identified in this analysis.
33 Asterisks indicate proteins with multiple subunits. Data indicate non-redundant peptide hits.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1. Patient clinical information

Sample	Group	Age	Sex	Race	Disease Type	ILD/ PAH	Disease Duration (years)	ANA Pattern	ANA Titer	MRSS
SSc 1	TOP1	36	F	White	Diffuse	mild ILD	2.5		1:320	43
SSc 132	TOP1	49	F	White	Diffuse	No		Homogeneous	1:640	27
SSc 218	TOP1	55	F	White	Diffuse	ILD		Homogeneous/Nucleolar	1:2560	18
SSc 208	TOP1	64	M	White	Diffuse	No		Nucleolar	1:1280	37
SSc 5	RNAP3	53	M	White	Diffuse	No	0.75	Speckled	1:80	36
SSc 7	RNAP3	45	F	Black	Diffuse	No	0.5	Speckled	1:80	27
SSc 10	RNAP3	52	M	White	Diffuse	No	0.5		0	22
SSc 18	RNAP3	69	F	White	Diffuse	ILD	0.5	Nucleolar	1:160	44
SSc 159	CENP	54	F	Mixed	Limited	No	7	Centromere	1:1280	2
SSc 177	CENP	64	F	White	Limited	No	15	Discrete Speckled	4+	
SSc 194	CENP	66	F	White	Limited	No	18	Discrete Speckled	4+	6
SSc 238	CENP	53	F	White	Limited	No	6	Centromere	1:640	5
SSc 226	CENP	55	F	Asian	Diffuse	No		Centromere	1:1280	6
HC 162	Control	24	M	White						
HC 400	Control	21	M	White						
HC 117	Control		M							
HC 118	Control		M							

				Prevalence in this dataset (avg/freq)				
Alias	Associated Proteins*	Disease Subset	Clinical Associations	Control (n = 4)	RNAP3 (n = 4)	TOP1 (n = 4)	CENP (n = 5)	Reference
Major autoantibodies								
RNA Pol III	POLR3A	dSSc	renal crisis, cancer	-	+++ (28/4)	-	-	Graf, et al. 2012; Mehra, et al. 2013
Scl70	TOP1	dSSc	poor prognosis, internal organ involvement, and proteinuria	+ (3/2)	+ (4/4)	+++ (19/4)	-	Mehra, et al. 2013
Centromere	CENPB, CENPH	ISSE/CREST	PAH, ILD	-	-	-	+ (1/2)	Mehra, et al. 2013
Other SSc autoantibodies present in our dataset								
Endothelial Cell	TUBB, VCL, LMNA, RPLP0	SSc	PAH	+ (1/1)	++ (6/4)	+ (4/2)	+ (0/1)	Dib, et al. 2012; Naniwa, et al. 2007
Fibroblast	ENO1, G6PD, HSPA1A, HSPA1B, VIM	SSc	PAH	+ (3/4)	+++ (12/4)	++ (5/3)	++ (8/5)	Terrier, et al. 2008, 2010
Histone	H1FX, HIST1H1B, HIST1H4A	SSc	PF, internal organ involvement, decreased survival	+ (1/1)	+ (3/3)	+ (1/1)	-	Mehra, et al. 2013
B23	NPM1	dSSc, CENP+ ISSc	PAH	+ (4/4)	++ (7/4)	++ (5/4)	++ (6/5)	Mehra, et al. 2013
Ku	XRCC5, XRCC6	ISSc	Myositis	+ (3/3)	+++ (12/4)	++ (8/4)	+ (2/3)	Graf, et al. 2012; Mehra, et al. 2013
Su	AGO2	SSc, PM/Scl	Unknown	-	+ (1/2)	+ (3/1)	-	Satoh, et al. 2013
Mitochondrial (M2)	DLD, PDHB	ISSc	Strong association with primary biliary cirrhosis	+ (1/1)	+ (2/3)	+ (1/1)	-	Mehra, et al. 2013
Pm/Scl	EXOSC1-10	SSc	PF, digital ulcers; decreased risk of PAH and GI symptoms	+ (2/2)	++ (5/3)	+ (2/2)	-	Mehra, et al. 2013
hnRNPs	HNRNPA1-3, HNRNP L	SSc	Common in SARDs	+ (0/1)	++ (7/4)	+ (3/4)	+ (2/4)	Siapka, et al. 2007
U1	SNRNP A, SPRNP70	SSc	Co-occurrence with SS-A/SS-B, PAH, overlap syndrome	-	+ (2/4)	+ (1/2)	+ (0/1)	Graf, et al. 2012; Mehra, et al. 2013
U5	SNRNP200	SSc, PM/Scl	Unknown	++ (6/3)	++ (9/4)	++ (8/3)	+ (1/2)	Kubo, et al. 2002
RO52/TRIM21	TRIM21	SSc	ILD, other autoimmune diseases	++ (6/3)	+++ (12/4)	++ (6/4)	++ (8/4)	Mehra, et al. 2013
RuvB	RUVBL1, RUVBL2	dSSc	Common in SARDs, older age at onset, male sex	+ (1/1)	++ (7/4)	+ (3/4)	+ (2/4)	Kaji, et al. 2014
Annexin V	ANXA5	dSSc, CENP+ ISSc	Digital ischemia	+ (2/2)	++ (7/4)	+ (4/3)	+ (3/4)	Mehra, et al. 2013
SS-B/LA	SS-A, SS-B	SSc	ILD, other autoimmune diseases	-	+ (3/4)	+ (2/2)	+ (0/1)	Mehra, et al. 2013
Peroxiredoxin	PRDX1	SSc	Disease duration, PF, cardiac involvement, TOP1+ patients	+ (2/4)	++ (8/4)	+ (3/3)	+ (4/4)	Mehra, et al. 2013
hUBF/NOP90	UBTF	ISSc	mild organ involvement, favorable prognosis	-	+ (1/2)	-	-	Mehra, et al. 2013
Th/To	POP1	ISSc	PF, renal crisis, poor prognosis, myositis, PAH	+ (1/2)	+ (1/1)	+ (3/3)	-	Graf, et al. 2012; Mehra, et al. 2013
PL-12	AARS	SSc, PM/DM	ILD without myositis	-	-	+ (1/1)	+ (1/1)	Hamaguchi, et al. 2013
OJ	IARS	SSc, PM/DM	ILD without myositis	+ (1/1)	-	+ (3/3)	-	Hamaguchi, et al. 2013
EJ	GARS	SSc, PM/DM	ILD, myositis	-	+ (3/4)	+ (2/2)	+ (0/1)	Hamaguchi, et al. 2013
Jo-1	HARS	SSc, PM/DM	ILD, myositis	-	+ (1/4)	-	-	Hamaguchi, et al. 2013
PL-7	TARS	SSc, PM/DM	ILD, myositis	+ (2/2)	++ (8/4)	++ (6/4)	+ (2/4)	Hamaguchi, et al. 2013
Ha	YARS	SSc, PM/DM	Interstitial pneumonia	-	+ (0/1)	-	-	Hashish, et al. 2005
Zo	FARSA, FARSB	SSc, PM/Scl	anti-synthetase syndrome	-	+ (2/4)	-	-	Betteridge, et al. 2007
SSc autoantibodies not detected in our dataset								
Fibrillarin	U3RNP	dSSc	More frequent in blacks; severe disease, poor prognosis	-	-	-	-	Mehra, et al. 2013
U11/U12 RNP	SNRNP35	SSc	Lung fibrosis, gastrointestinal involvement	-	-	-	-	Mimori, 1999
PDGFR	PDGFR	SSc	Unknown	-	-	-	-	Svegliati Baroni, et al. 2006
MMP	MMP family	dSSc	Skin, lung, and vascular fibrosis	-	-	-	-	Mehra, et al. 2013
tPA	PLAT	ISSc	PAH	-	-	-	-	Mehra, et al. 2013
IFI16	IFI16	ISSc	Common in SARDs	-	-	-	-	Mehra, et al. 2013
Fibrillin 1	FBN1	dSSc	Choctaw and Japanese patients; absent in Caucasians	-	-	-	-	Mehra, et al. 2013
Vascular Receptors	AGTR2, EDN1	SSc	TOP1+ patients, renal crisis	-	-	-	-	Mehra, et al. 2013
ATF2	ATF2	SSc	Longer disease duration, decreased lung function	-	-	-	-	Mehra, et al. 2013

Figure 1. Overview of Mass Spectrometry Results



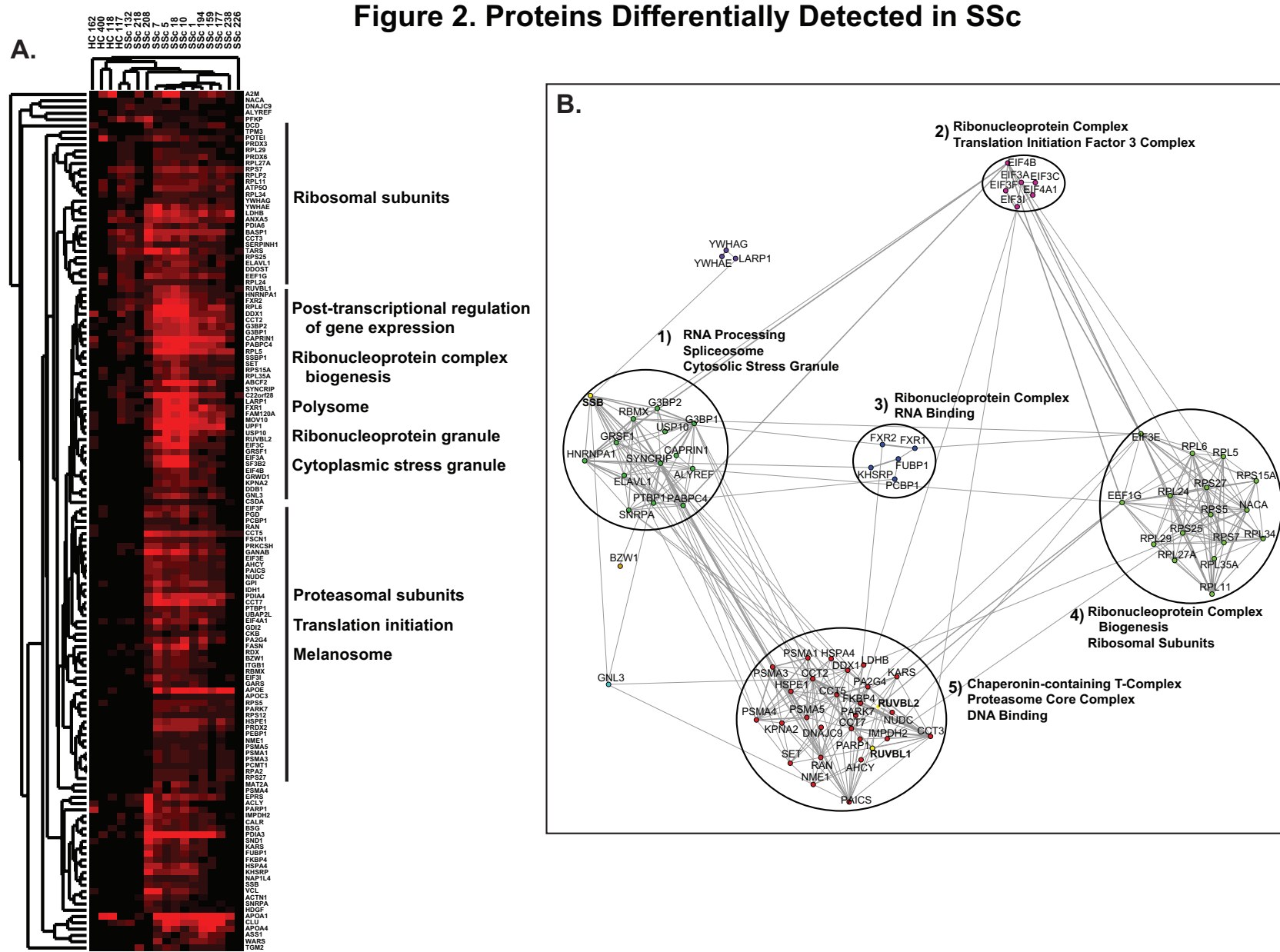


Figure 3. Validation of RNA processing bodies and stress granules as targets of the SSc autoantibody response

